File 343
acc. 2246

From Thumbnail to Discourse

Notes from the Nasuli Syntax Workshop
16 July - 31 August 1979

Jan Forster
Austin Hale
Ken Maryott

The 1979 Nasuli Syntax Workshop was originally conceived of as a workshop on low-level syntax with the primary goal of assisting the participants in the preparation of brief grammatical sketches which would do three things: 1) concisely summarize what was known about grammatical systems such as noun phrase, verb phrase, clause, and sentence in the various languages under study; 2) state the current hypotheses in a way that would invite testing against new data; and 3) provide an outline that would be modified and expanded by periodic up-dating so as to serve as an ongoing loose leaf reference grammar of the language under study. The success of the workshop was not to be measured in terms of published papers that resulted from it, but rather in terms of the effect it had in increasing the payoff experienced by the participants from the time invested in linguistic analysis.

In looking at the list of participants who planned to take part in the workshop, the workshop staff felt it appropriate to attempt to stimulate research far beyond the normal bounds of low-level syntax. While we were prepared to accept solid, hierarchically-oriented descriptions of phrase and clause as the basic bread-and-butter goal of the workshop, we were also prepared to extend our goals in two directions: 1) that of extensive semantic (referential) indexing of the descriptions in terms of categories that cut across hierarchical levels, admitting of variant forms of expression on several different levels of structure; and 2) that of investigating the relationships that link high-level discourse systems to choices between various options in low-level grammar. The first of these directions aims

at answering the question "What are all the ways in which ___ can be expressed in the language under study?" The second aims at answering the question "What determines the 'free' options and the 'peripheral' choices available at phrase, clause, and sentence levels?"

What follows are the notes which accompanied the lectures at the workshop. It is our hope that these notes are sufficiently full to enable someone who was not in attendance to benefit by reading them.

1. Orientation Lecture

A. Goals of the Workshop.

The workshop has three listed priorities. The first of these is thumbnail sketch coverage of lower-level grammar. This priority was clearly articulated by Evan Antworth, linguistic chairman in the letter that set the workshop in motion: "I think it's fine to include discourse analysis in the workshop as long as the participants first have a solid foundation in lower level grammar ... the workshop will aim for thumbnail-sketch-type coverage rather than writing formal papers." As we interpret this it will involve three kinds of activities: a) hierarchical description of norms and observed variants for the structure of noun phrase, verb phrase, and, as time permits, clause and sentence; b) an enumeration of options available within these structures which are not controlled within the structures themselves but rather are controlled by higher level systems; and c) sharing of results and problems by each participant with all other participants in the group both in written form and by way of oral group presentations. The form of presentation is to be one that could function as a loose leaf grammar filing system for future language study. A good hierarchical breakdown of a language is analogous to the alphabetical order in a dictionary. It constitutes an excellent filing system if properly utilized. The workshop staff will need two copies of all last drafts before the end of the workshop.

The second priority is that of indexing the thumbnail sketch for semantic (referential) categories that cut across the hierarchical description. This will involve answering the question, "What are all the ways in which ___ can be expressed

in the language under study?" The index will provide semantic access to the grammar and can be compared to the English index or thesaurus associated with a dictionary. Here again participants will be expected to share their problems and results in both written and oral presentations with all other members of the workshop.

The third priority is that of identifying the higher-level choice systems that influence the exercise of lower-level options. This will involve tying the lower-level grammatical structure of the language under study into the discourse structure in all its many aspects. Again, sharing of results and problems in written and oral presentations is expected.

B. Rationnale for Thumbnail Sketches.

Those who get the most done in linguistic research typically oscillate between intensive work on detailed analysis on the one hand and a broad overview of grammatical structure on the other. If one concentrates exclusively upon the details of analysis he is likely to lose his way and become bogged down in a morass of tiny facts. If one looks only at the over-all structure of the language, he is not likely to master any of it in detail. It is our impression, however, that most of us are more in danger of being swamped by detail than we are of avoiding it altogether. For this reason we are recommending a loose leaf approach to over-all grammatical structure as a supplement to other more detailed approaches to the analysis of particular grammatical systems.

The general guidelines for the construction of a thumbnail sketch can be summarized in the following ten recommendations.

1. The first 85% of what one will ever explicitly analyze of the grammar or phonology can often be dealt with quite early in ones study of a language. The remaining 15% will take you the rest of your life. Shoot first for the most accessible 85%.

2. Lead with hypotheses. Don't let data accumulate. Keep your loose leaf grammar organized in a topical fashion and keep it up-to-date with key hypotheses and data in summary form. Keep hypotheses up-to-date with the data.

3. For a rapid first sketch keep your focus on the surface and try to select a formal apparatus that relates simply and directly to the surface. Aim first at observational adequacy.

4. Select a few topics and exhaust your data in the description of these topics (but shallowly at first). As you get more and more data, your analysis should require progressively less and less revision to accomodate the facts if you stick to the surface.

5. As soon as your hypotheses in a given area of the grammar begin to account for most of the data, move on to look at another area. Keep your survey topics small enough and shallow enough to allow you to move rapidly toward surface coverage of all major structures both in phonology and in grammar. Do what is easy first. What is out of reach may become easy after you have done a number of systems that are easy and accessible in terms of where you are when you attack them.

6. List at least a dozen examples for each hypothesis you enter in your sketch.

7. Be sure to list your assumptions, queries, uncertainties for later follow-up. You will need to do detailed follow-up sometime and you do want to oscillate between detailed specific studies and the less detailed over-all view. Put down your unconfirmed hunches but be sure to label them as such.

8. Date all major entries so that you can reconstruct what your over-all view was at a given point in time, and so that you will know whether sub-system A has been reworked in the light of findings within sub-system B or not.

9. Refuse to get bogged down in detail when you are operating in the thumbnail sketch mode of analysis.

10. Keep your write-ups as short and concise as possible. Do not let your thumbnail grow to encyclopedic proportions. It is not your master file for data. It is a means for maintaining an overview of the structure, a way of keeping your key hypotheses, rules, and systems maximally accessible for checking against new data. It is a very large-scale map. It should contain references to masses of data, but limit the cited examples to those which are most cogent and convincing. It is a good place to record alternative hypotheses regarding the larger over-all systems.

C. Some Basic Working Concepts.

A thumbnail sketch is supposed to be a device or strategy for maintaining a portable, up-to-date summary of the over-all structure of the language under study. It should make key hypotheses easily available for checking against data. The formal details of such a sketch are a matter of personal preference (even though we will be making lots of rather specific suggestions in these lectures) and although few linguistic theories are content to stop with a description of surface grammar, they all must deal with the surface at some point in the description, and are thus eligible as frameworks for a thumbnail sketch. The set of basic concepts presented here is one that has been found useful, but is by no means the only set that could be fruitfully proposed for this purpose.

This section is only a beginning. It touches only a few concepts and does so only briefly. We hope that the fuzziness which arises here from the brevity of its discussion will wear off as the concepts are put to work in later lectures.

1. Hierarchy and the Outline of a Thumbnail.

For any kind of reference grammar or loose leaf grammar filing system, the usefulness of the item depends in large measure upon the outline. A good outline should leave one in no doubt as to where to look for answers to a given question, and it should likewise leave one in no doubt as to where some new insight into the grammar should be recorded. Furthermore, such an outline should be inherently stable. Ideally, changes in one section of the grammar should not result in wholesale modification of the outline or in extensive modifications to other sections of the grammar. The grammar should be modular, allowing one to modify, MODULAR or even interchange, components with minimal disruption of the rest of the system.

[[The cost of such stability is a certain level of redundancy, the kind of redundancy that led Chomsky and Halle to reject the taxonomic phonemic level as a well-motivated level of representation in a generative grammar. The kind of modularity that is here viewed as essential will require that certain rules be stated twice. Our feeling is that for our present purposes such a price is a

very small one to pay for modular stability.]]

One can go about the construction of an outline for a grammatical sketch in various ways. First of all, we would like to distinguish three distinct starting points for the description of language:[1] the sound system (phonology), the grammar, and the semantics (or referential hierarchy). Almost everything in a given text can be analyzed in three ways: by looking at its phonological realization, its pronunciation and the role it plays phonologically within the total text; by looking at its grammatical structure and the role it plays within the grammatical structure of the total text; and by looking at its semantic content and the contribution of that element to the interpretation of the message of the total text. A thumbnail sketch could logically be divided into three sections, phonology, syntax, and semantics. Since phonology does not fall within the scope of this workshop, it will not be developed further in these lectures. Our first priority will be in the area of syntax, with some regard for semantics.

Consider, then, how one might construct an outline for the grammatical structure of a language. One time-honored strategy for achieving a certain degree of modular stability is that of dividing a language up into separate levels, and within each level, into various types of units. The basic question of hierarchy that we wish to consider here is, "How can the part-whole relations within grammatical units best be utilized in organizing a thumbnail sketch?"

Immediate-constituent analysis[2] represents an early approach to syntactic

---

[1]We will be drawing freely from the view of language presented by K.L. Pike and E.G. Pike, 1977, Grammatical analysis (SIL Publications in Linguistics and Related Fields, Nr. 53) Dallas: SIL-UTA. There these three starting points are hierarchies.

[2]For various discussions of this approach see: Rulon Wells, 1947, Immediate constituents, Language 23:81-117 [Reprinted in Joos, 1957, Readings in linguistics Washington, D.C.: ACLS, pp. 186-207]; Richard S. Pittman, 1948, Nuclear structures in linguistics, Language 24:287-292 [Also in Joos, pp. 275-278]; Charles F. Hockett, 1958, A course in modern linguistics, New York: The Macmillan Co., Ch. 17. For an extended description of English in these terms, see Eugene A. Nida, 1960, A synopsis of English syntax (SIL Publications in Linguistics and Related Fields, Nr. 4) Norman: SIL. For a description of a non-Indo-European language in these terms see Robert B. Jones, Jr., 1961, Karen linguistic studies, description, comparison, and texts (University of California Publications in Linguistics, Vol. 25) Berkeley and Los Angeles: University of California Press.

analysis in which some value was placed upon making successive, often binary, cuts in an utterance until all segmentable parts had been analyzed out. The first cuts were those made between major, high-level, relatively independent parts of the utterance. Thus, a sentence such as _the teacher saw John's book_ might first be cut into two parts, _the teacher_, and _saw John's book_. The second cut might come between _saw_ and _John's book_. A third cut might separate _the_ from _teacher_ and _John's_ from _book_.

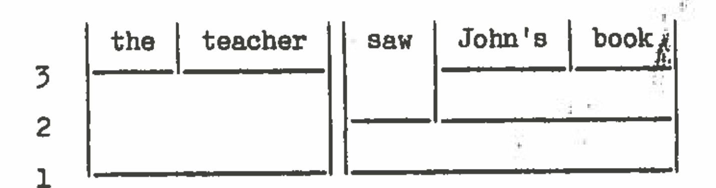


Figure 1. Successive binary cuts in an IC-analysis.

Further cuts could be made to separate the agentive derivational affix, _-er_, from _teach_, and to separate the possessive suffix, _-'s_, from _John_, and conceivably even to separate the past tense morpheme from the verb _saw_. This approach was used by Hockett, Nida, Jones, and others to account for a very broad range of syntactic structures. It was later utilized by Chomsky[3] and others in a somewhat modified form(called phrase structure grammar) to generate a small subset of structures which were at first called 'kernel sentences' and later on 'deep structures'. In the 1965 version of the theory, the phrase structure grammar generated trees analogous to the IC-analysis presented above which were then interpreted by the semantic component. Before such a 'deep structure' representation could serve as an account of surface structure, however, it had to pass through a series of transformational rules. The interesting thing to notice here is that although the deep structures tended to be highly layered and were often binary, much the way the representations of IC analysis tended to be, the effect of the transformational rules was generally to reduce the layering and increase the average number of

---

[3]Noam Chomsky, 1957, _Syntactic structures_ (Janua Linguarum Nr. 4) The Hague: Mouton and Co., Chapters 4 and 5. For the 1965 version see Chomsky, 1965, _Aspects of the theory of syntax_, Cambridge, Mass.: MIT Press.

branches eminating from any given node in the tree. In short, while deep structure tended toward IC analysis (more layering, less branching), surface structure tended in the direction of string-constituent analysis (less layering, more branching from any given node).[4]

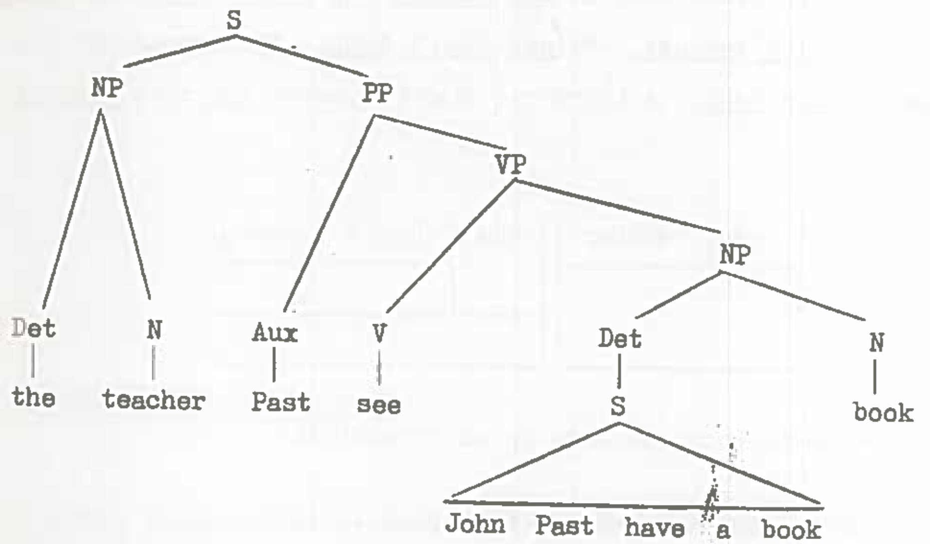


Figure 2. Simplified 'deep structure' representation.

While other approaches were concerned with making sense of abstract underlying representations and their semantic interpretations, adherents of Tagmemics continued to be concerned with making sense of surface structure. One important step for our present purposes was the development of the syntagmeme, or later, the root as applied to clause and sentence structure.[5] Instead of numbering successive cuts, or labelling brackets at whatever level in an underlying tree they might occur, here was an effort to systematize the layering of constructions within surface structure. Longacre mentions this as one of the fundamental insights of Tagmemics.[6]

[4] For an insightful discussion of this see Robert B. Lees, 1964, Review of Zellig S. Harris, String analysis of sentence structure, IJAL 30:415-420.

[5] For treatments of clause and sentence roots and stems see: Evelyn G. Pike, 1974, Coordination and its implications for roots and stems of sentence and clause (PdR Press Publications in Tagmemics Nr. 1) Lisse, Netherlands: Peter de Ridder Press, as well as Pike and Pike, 1977, Grammatical analysis, pp. 12, 21-26, 39ff., 145, 262ff.

[6] Robert E. Longacre, 1965, Some fundamental insights of tagmemics, Language 41:65-76. See also: Peter H. Fries, 1971, Some fundamental insights of tagmemics revisited, in Polome, Winter, and Jazayery (Eds.), In honor of the retirement of A.A. Hill from The University of Texas.

As we currently see it, this pressure to systematize the hierarchical organization of surface structure culminates in the notion of paired levels in the grammatical hierarchy.[7] In this development, levels are systematized not only in terms of the internal structure of syntagmemes, but also in terms of the various functional thresholds which exist within the grammatical hierarchy. Not all languages have a clear-cut structural contrast between word and phrase, or between sentence and paragraph, but all languages presumably have a clear functional distinction between a structure that names a term and a structure that asserts a proposition. It is possible to systematize our description of levels in the grammatical hierarchy according to the various functional thresholds which can be thought of as the 'meanings' of the various levels within grammar.



Meaning	Minimum Unit	Expanded Unit
Social Interaction	Exchange	Conversation
Theme-Development	Paragraph / Sentence Cluster	Monolog
Proposition	Clause	Sentence
Term	Word	Phrase
Lexical Package	Morpheme	Morpheme Cluster

Figure 3. Paired Grammatical Levels (from Pike and Pike, 1977, page 24)

From this point of view it is plausible to attribute the following kind of organization to the surface structure of our sample sentence (See Figure 4).

The three major thresholds in Figure 3 are separated by double lines. Term and Lexical package have to do with the units that refer to participants, props, and various other entities. They also have to do with units that refer to actions and states that are predicated of or attributed to entities, participants, and

---

[7] We part company with Longacre and follow Pike at this point. See Robert E. Longacre, 1976, An anatomy of speech notions (PdR Press Publications in Tagmemics Nr. 3) Lisse: The Peter de Ridder Press, pp. 284-286. For an interesting discussion of thresholds see K.L. Pike, Thresholdism versus reductionism to appear in Seyles (ed.) For Hansjakob Seiler (Tübingen: Verlag Gunter Narr) pp. 53-58.

and props. At the level of the Term and below we are concerned with describing the internal structure of these naming and referring units. The next major threshold above this includes Theme-Development and Proposition. This threshold has to do with units that make assertions, ask questions, give commands, and that develop themes or topics at greater length. It is at this threshold that we are concerned with how to say something about something.

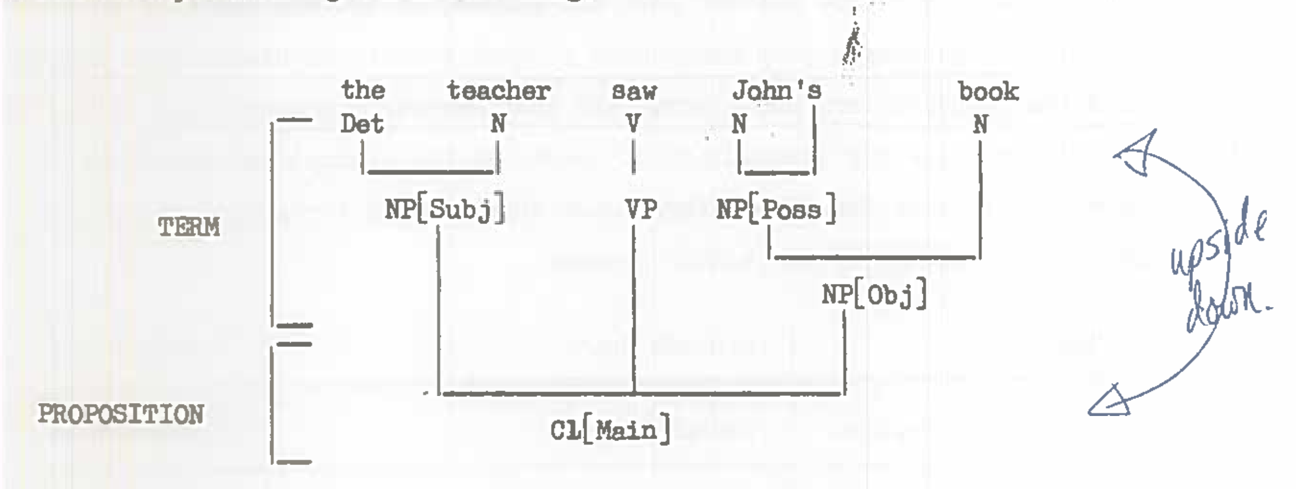


Figure 4. Sample Thumbnail Tree with functional thresholds indicated.

At the level of Development and below we are concerned with describing the internal structure of these propositional and developmental units. The last major threshold is that of Social Interaction. This threshold has to do with verbal (and behavioral) interaction between at least two parties to a conversation. At the level of Interaction we are concerned with describing the strategies by which two or more speakers interact to achieve various ends.

From this point of view we would suggest that a thumbnail sketch of any human language can start with three major headings: 1) Phonology, 2) Grammar, 3) Semantics. Furthermore, within Grammar it will have at the very least a) Term, b) Development, and c) Exchange. Optionally it will also have as much of the structure of Figure 3 as is appropriate for the language under study.

We have asked the question, "How can the part-whole relations within grammatical units best be utilized in organizing a thumbnail sketch?" So far we have expressed a preference for an outline in which the grammatical units are sorted out according to the various functional thresholds that they cross. We have also expressed a preference for the string-constituent trees that appear most appropriate to the

organization of the surface structure rather than for the more heavily layered and more nearly binary trees that are more appropriate to the organization of underlying structure. We have proposed a small number of very general headings for the outline. We have indicated a rather strong preference, for reasons of modular stability, for a description that makes use of a system of levels and types of unit, rather than one which consists of a monolithic set of integrated rules. An important part of answering the question posed at the beginning of this paragraph is to show, in terms of some specific language, just how the more specific headings in the outline are determined. This will be attempted in the following lecture. Rather than pursue the question of outline development further at this time, we turn to another question that needs answering, namely, "What is the nature of a grammatical unit within a thumbnail?"

2. The Nature of Grammatical Units in a Thumbnail.

It seems to make good sense to agree with Pike and others who admit to the relevance of units in a linguistic description. A well-delineated unit is described in terms of 1) contrastive features that distinguish it from other units, 2) the range of variation within which it maintains its identity, and 3) its range of membership in various classes of units, its range of occurrence in various syntagmatic sequences, and its range of relationships to various systems.[8] The sample thumbnail tree in Figure 4 may not look much like the tagmemic trees of Pike and Pike, but in the following section we will attempt to show that they are closely related. The main labels on the various nodes (or horizontal lines) of Figure 4 (Det, N, NP, VP, CL) are syntactic category labels which serve to name various construction types and parts of speech. These labels belong to Cell 2 of the four-feature grammatical tagmeme. The features from the other cells are entered as needed or desired within square brackets. Thus we might have labeled the teacher

---

[8] For a recent discussion of this see K.L. Pike, Here we stand--creative observers of language (presentation at the Colloquium on Language Development [Child language acquisition] in Paris in connection with his honorary doctorate, December, 1978). See also the first chapter of Pike and Pike, 1977, Grammatical analysis.

in Figure 4 as NP[Subj, Act, Human, Sg]. The feature [Subj] would have been from Cell 1, [Act] from Cell 3, and [Human, Sg] from Cell 4. One certainly could use the four-cell diagram. For ease of typing and for the saving of space, however, we will often use a reduced version of the four-feature tagmeme.

Cell 1 features		Cell 1: SLOT	Cell 2: CLASS		Cell 2 features	
Subj, Obj, Pred, Compl	Head, Poss	> > > > SLOT (1)	CLASS (2) < < <		NP, VP, CL, S	Det, N, V, Adj
Cell 3 features		**Cell 3: ROLE**	**Cell 4: COHESION**		**Cell 4 features**	
Item, Act, Und, Site		> > > > ROLE (3)	COHESION (4) < <		Sg, Pl, Fem, Masc	

Figure 5. The four-feature grammatical tagmeme (Pike and Pike, 1977, Page 35).

Each of the four cells in Figure 5 is important. The features of Cell 1 are grammatical relations having to do with the organization of material in prominence and attention, and are often closely tied to grammatical markings of linguistic forms. Frequently used features of Cell 1 include [Subj] 'subject-of', [Obj] 'object-of', [Pred] 'predicate-of', [Compl] 'complement-of', [Head] 'head-of', and [Poss] 'possessor-of'.

The features of Cell 3 are semantic relations linked to the syntactic structure. Frequently used features of Cell 3 include [Act] 'actor-of', [Und] 'undergoer-of', [Site] 'site-of' (or [Sco] 'scope-of') [Item] 'item-of', and, where such relations require more detailed discrimination, case labels may be used instead.

The features of Cell 4 are cohesive relations, typically involving agreement patterns relating to tense, number, gender, location, honorifics, temporal sequence, narrative sequence, and the like.

The sample feature sets given as illustrations here are largely drawn from clause level, but there are analogous sets for all the other levels. For further discussion see Pike and Pike, 1977, Chapter 3 (pp. 35-68) and Appendix 3 (pp. 455-467). Appendix 3 gives an etic list of features for all four cells for all levels of the hierarchy. We will not necessarily restrict ourselves to the features listed there, but it is a good place to look in order to get a feel for what belongs in each of the four cells.

3. The Use of Norms to Simplify the Representation of Units.

One of the apparent regularities of language that makes it possible to use a reduced representation of the grammatical four-cell tagmeme is that for peripheral tagmemes of verious constructions the features for the four cells are often either identical or so closely related as to be predictable one from the other. For nuclear tagmemes there are often norms which relate the features of Cells 1 and 3 to one another. The reduced representation allows us to take note of deviations from the norms when these occur while omitting features from various cells when these can be predicted on the basis of the norms. In all cases we take the class or category as the name of the unit, and features from other cells are viewed as specifying relevant grammatical relationships of the unit to its context. Relationship features thus appear in square brackets appended to unit names.

Consider the following sentences as an illustration of the skewing among cells in the nucleus of clause and the relative redundancy among cells as one moves from the nucleus toward the margin or periphery.

a. Last Wednesday John felled the tree with an axe.

b. Last Wednesday the tree was felled by John with an axe.

c. Last Wednesday an axe felled the tree.

d. Last Wednesday was the day the tree was felled.

In Examples (a), (b), and (c), Last Wednesday could have the following kind of four-cell representation. (NP = noun phrase, Mar = margin)

Mar	NP
Time	

> NP[TIME]

A simplifying norm which could be relevant here is that Time is normally part of the grammatical periphery of a clause. From the role, Time, we can predict the slot, Mar. (Given only the slot, Mar, however, we cannot predict the role, Time, since Time is but one of many roles to occur within the margin.) Such a norm needs to be made explicit by recording it in a list of assumed norms, and having done that we may abbreviate the four-cell representation given above as simply NP[Time] whenever it occurs. Whenever the norm does not apply, the full set of features need to be given.

In Example (d), Last Wednesday was the day the tree was felled, the phrase, Last Wednesday is drawn into the clause nucleus and has a rather different four-cell representation.

Subj	NP
Item	

Here the grammatical slot filled by Last Wednesday is that of subject and the role is that of Item. It has time as lexical content, but not as role within the clause. In the nucleus, especially at an early stage of analysis, we would be more inclined to include all the features: NP[Subj, Item]. It may turn out, however, that another simplifying norm will emerge, namely, that the subject of a copular verb is normally an Item. Once this norm has shown its worth and has been properly listed among our assumed norms, we may write simply NP[Subj]. Once such a simplifying norm has been accepted, however, instances that violate the norm must be represented in a way that makes their deviation from the norm explicit.

In Example (a), Last Wednesday John felled the tree with an axe, the noun John, fills the subject slot and has the role of actor.

Subj	NP
Act	

If we take as our norm for English that in a transitive clause the subject is actor, we may elect simply to write NP[Subj] in this instance.

In Example (b), Last Wednesday the tree was felled by John with an axe, the noun phrase, the tree, fills the subject slot but has the role of undergoer.

Subj	NP
Und	

If we have taken the norm to be that subjects of transitive clauses are actors, we are under pressure here to record the deviation from the norm and not to simplify: NP[Subj, Und].

In Example (c), Last Wednesday an axe felled the tree, the noun phrase, an axe, fills the subject slot. If it is here viewed as having the role of instrument, this would also represent a departure from the assumed norm, and would need to be represented. The four-cell tagmeme,

Subj	NP
Inst	

would be represented as NP[Subj, Inst] in this instance.

In Examples (a) and (b) the noun phrase, with an axe, fills an adjunct (Adjn) slot and has the role of instrument. The four-cell representation would be:

Adjn	NP
Inst	

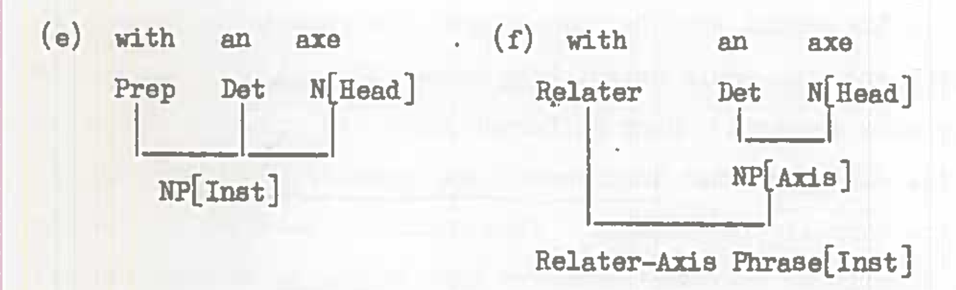
If we take adjunct as the normal slot for non-subject instruments in English we may simply write NP[Inst]. We would select Inst rather than Adjn as the relevant feature since it is more specific. Many different roles can occur in the grammatical slot of adjunct. The norm says that instruments are normally adjuncts. It does not say that adjuncts are normally instruments. This choice is parallel to the choice of NP[Time] as the simplified representation for Last Wednesday in Examples (a), (b), and (c). Given that Last Wednesday has the role of Time, the norm says that the slot is Mar. Again, margin is a slot in which many different roles can occur. From our statement of the norm, the slot feature, Mar, is predictable given the role feature, Time, but given only Mar, no specific role feature can be predicted.

It would seem, then, that in the less nuclear slots the role is preserved in our simplified representation and the slot is supplied by norm wherever possible. In nuclear slots, such as Subj, it may be useful in some languages to set up a norm to predict the role from the slot. In other languages it may be more useful to predict the slot in terms of the role. In either case we will often end up with a full set of features in our representation of the nuclear units because it is in the nuclear units that skewing and departures from norms for correlations between slot, class, and role are most frequent.

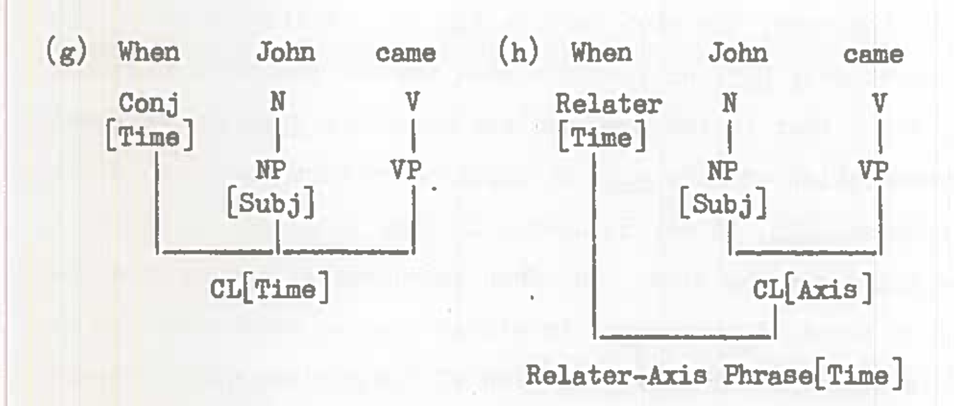
The line between Margin & Adjunct is not apparent on the surface. An adjunct can be a margin. Both are very general terms.

4. Layers and Strings in Surface Trees.

When one is involved in constructing a surface tree for a whole monologue, it is often advantageous to reduce layering whenever the reasons for maintaining if are not strong. Unconstrained layering may be advantageous for abstract underlying representations, but it does tend to make surface trees difficult to work with. It also tends to multiply construction labels and construction types, often without much payoff. One general policy that has been advantageously followed is that of grouping relators as co-constituents of the unit related rather than separating the axis and the relater as co-constituents of a higher-layered construction. In the absence of strong reasons to the contrary, Tree (e) is to be preferred to Tree (f) for most purposes in a thumbnail. [Tree (e) could even be considered an abbreviated form of Tree (f) for those who wish to retain Tree (f) as the full official representation of structure.]



Likewise, in the absence of strong reasons to the contrary, Tree (g) is to be preferred to Tree (h).



In the case of noun phrases this policy allows noun phrases to have a complete paradigm of case forms even where some are zero marked, some are marked by case affixes, and others are marked with prepositions or postpositions. The minor

complexity of the realization of case marking is thus contained on the lower level, allowing a somewhat tidier picture of case relations at the next higher level. It also allows us to make sense of categories such as locative pronoun. If in the house is viewed as a noun phrase, then there can be a locative pronoun in a rather straightforward sense. In many languages this view allows a better parallelism of nominal and pronominal forms than one which makes major use of relater-axis constituent structure in the surface trees.

Another issue that usually arises in the construction of surface trees has to do with the fact that a unit can be grammatically bound on one level but have a function on a very different level.

(i) John likes tomatoes.

In Example (i) the present singular suffix, -s, on the verb, like, is clearly grammatically bound on the word level. The tense that it signals, however, holds for the whole clause, as does the singular-subject cross reference. The policy adopted here for thumbnail surface trees is the following: When the level at which a unit is grammatically bound is different from the level or levels on which the unit functions semantically, let the surface tree represent the relation that the unit bears to the level at which it is grammatically bound. Some means other than that of surface trees will be required in order to indicate functional semantic relations that skew with the grammatical ones. One way of doing this will be discussed in later lectures on semantic indexing.

D. Suggested Reading.

In addition to reading the items that have been referred to in the footnotes thus far, the following are recommended as important supplements to these notes.

Pike, K. L. 1975. On describing languages. (PdR Press Publications in Tagmemics - 2) Lisse: The Peter de Ridder Press.

Thomas, David. 1975. Notes and queries on language analysis. (Language Data, Asian-Pacific Series, Nr. 10) Huntington Beech: SIL.

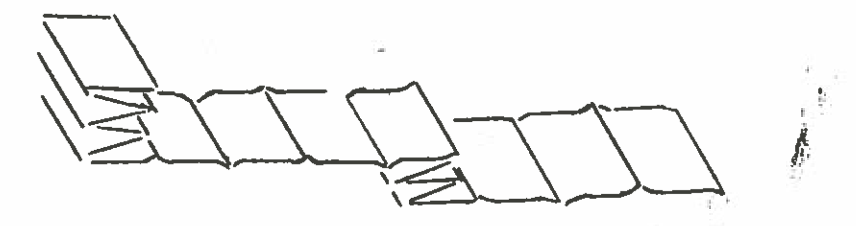
Welmers, William E. 1975. Data for a grammatical outline, in Thomas, 1975, pp. 103-112.

2. Development of a Thumbnail from Text to Outline

The purpose of this lecture is to present an overview of the steps in one well-tested approach to the construction of the grammatical portion of a thumbnail sketch. The major question approached will be that of constructing an outline that relates to the hierarchical grammatical structure of the language under study in such a way as to provide a natural filing system for grammatical insights, hypotheses, and descriptive statements. The process outlined here consists of six steps: A. The Text Accordian; B. The Text Tree; C. The Work Chart; D. Broader Coverage: Total Filing, Concordance Search; E. The Formal Summary; and F. The Outline. We shall discuss each of these briefly in turn.

A. The Text Accordian.

Step one is to lay out a text on sheets of lined paper in a three-line format. The sheets are glued together side by side and folded as an accordian. The accordian fold allows any desired stretch of text to be laid out for study. It also allows any two or three stretches to be laid out side by side for comparison. The entire text can be stored conveniently in a file folder.



The text is written in the source language across the top line of each page. Where possible it is broken into sentences which are numbered consecutively for ease of reference. On the next line down is a word-by-word (or better yet, morpheme-by-morpheme) translation of the text into English (or into the language in which the consultant and consultee communicate most easily). At the very bottom of the page is a sentence-by-sentence translation of the text. Figure 1 is a sample fragment of a text accordian.

Some comments on the arrangement of the material on these three lines. Regarding the two top lines, it is very helpful if the first letter of each source language

word is lined up directly above the first letter of its English gloss. Regarding the top line and the bottom line, it is very helpful if the sentences of the bottom line are numbered to correspond with the sentences of the top line. Furthermore, corresponding sentence numbers in the top and bottom lines should directly lined up, one above the other.

khicā-yā	macā-ta	1.	cha-guu	desa-e	cha-mha	mahārānii	du.	2.	wa	mahārānii-yā ...
dog-Gen	child-Pl		one-Cl	country-Loc	one-Cl	queen	is		that	queen-Gen

The children of the dog   1. In a certain country there was a certain queen.   2. As for that queen ...

Figure 1. Sample stretch of an accordian.

An accordian made from sheets of the intermediate pad size (7 1/2 by 10 inches) is about the minimum. An accordian made from lined foolscap or legal size paper is very nice to work with.

If a morpheme-by-morpheme translation is given and morpheme boundaries are given in the source language, the corresponding boundaries should be given in the line of glosses as they are in Figure 1. If morphemes are glossed but not segmented in the source language, commas may be used to separate morphemes in the line of glosses, as follows:

khicāyā	macāta	1.	chaguu	desae	chamha	mahārānii	du.	2.	wa	mahārāniiyā
dog,Gen	child,Pl		one,Cl	country,Loc	one,Cl	queen	is		that	queen,Gen

This kind of representation may make the text more difficult to use for morphology.

The accordian need not be typed, but if it is be sure to leave enough room for the hand-written tree and for the various comments that will be written in on the next step.

B. The Text Tree.

Step two is to assign a surface structure tree to the text. Constructing such a tree for a text involves a considerable amount of analysis and it will not be possible within the scope of one lecture to make explicit all that needs to be taken into account when such a tree is constructed. A few things, however, can profitably be mentioned. (1) Use a pencil. It allows you to change your mind without making a new accordian. (2) Don't try to decide everything the first time through. On the first pass through the text pass over really tough sentences in favor of the simpler ones that may help you set up surface patterns which will make the more difficult sentences easier to tackle. (3) When two or more analyses are possible for a given sentence or construction, give them both, one below the other, and come back again at a later time to sort them out. Keep a list either mentally or on paper of as many patterns as you can for each of the major construction types (NP, VP, CL, S). Valid patterns tend to recurr. Those that do not are suspect. Furthermore, each pattern you use in assigning a tree to a construction should have testable semantic consequences. A good tree is a potentially useful device for the construction of an English back-translation that both makes the meaning of the original clear in English and at the same time gives some fairly clear indication of the salient structures of the original. Though deep structures are a more consistent basis for semantic interpretation, surface structures do have consistent semantic correlates. Make use of these in controlling the analysis. From these comments it should be clear that constructing a discourse tree is not an ad-hoc sentence-by-sentence process. Every tree assigned to every construction is a hypothesis: first of all that the language has such a construction, and secondly that this string is an instance of that construction. Actually, to keep adequate track of the system at this stage it is quite useful to begin the formal summary very early in the process of treeing the text (See Section E. The Formal Summary). By working back and forth between the trees assigned to the text and the formal summary, the full impact of the interaction between text and the grammatical analysis can be profited from most beneficially. Furthermore, the fact that you are working with the full range of constructions from morpheme to discourse gives a balanced feel for the whole that is hard to gain in any other way.

B. The Text Tree.

Step two is to assign a surface structure tree to the text. Constructing such a tree for a text involves a considerable amount of analysis and it will not be possible within the scope of one lecture to make explicit all that needs to be taken into account when such a tree is constructed. A few things, however, can profitably be mentioned. Use a pencil. It allows you to change your mind without making a new accordian. Don't try to decide everything the first time through. On the first pass through the text pass over really tough sentences in favor of the simpler ones that may help you set up surface patterns which will make the more difficult sentences easier to tackle. When two or more analyses are possible for a given sentence or construction, give them both, one below the other, and come back again at a later time to sort them out. Keep a list either mentally or on paper of as many patterns as you can for each of the major construction types (NP, VP, CL, S). Valid patterns tend to recurr. Those that do not are suspect. Furthermore, each pattern you use in assigning a tree to a construction should have testable semantic consequences. A good tree is a potentially useful device for the construction of an English back-translation that both makes the meaning of the original clear in English and at the same time gives some fairly clear indication of the sálient structures of the original. Though deep structures are a more consistent basis for semantic interpretation, surface structures do have consistent semantic correlates. Make use of these in controlling the analysis. From these comments it should be clear that constructing a discourse tree is not an ad-hoc sentence-by-sentence process. Every tree assigned to every construction is a hypothesis: first of all that the language has such a construction, and secondly that this string is an instance of that construction. Actually, to keep adequate track of the system at this stage it is quite useful to begin the formal summary very early in the process of treeing the text (See Section E. The Formal Summary). By working back and forth between the trees assigned to the text and the formal summary, the full impact of the interaction between text and the grammatical analysis can be profited from most beneficially. Furthermore, the fact that you are working with the full range of constructions from morpheme to discourse gives a balanced feel for the whole that is hard to gain in any other way.

To begin with it is useful to take advantage of the threshold functions in the grammatical hierarchy as points of departure for treeing the text. Strings which function as Terms serve to refer to participants, props, situations, times, places, and the like. Strings which function as Propositions serve to say something about participants, props, situations and the like. Propositions relate props, participants, and situations to predicates of action or state.

It is possible to take an initial pass through the text in order to identify noun phrases and assign trees to them. There is a minor risk involved in this, however. When clause trees are assigned it may be necessary to revise the noun phrase trees in the light of clause structure.[1] This risk is minimized if one attempts to construct trees up to sentence level on the first pass. This also has the advantage of forcing an early confrontation with the various trade-offs that can be made among units. In the interests of modular stability it is very important to be aware of these trade-offs from very early on. The kind of structure one encounters from sentence on up is often sufficiently different from that of sentence and below that Hale, at least, prefers to save the higher levels for a later pass through the text.

Consider now a sample of the treeing process for a brief stretch of Newari text.[2] We will start with a story by Prem Bahadur Kansakar entitled 'The children of the dog'.[3]

khicā-yā	macā-ta
dog-Gen	child-Pl

. . .
The children of the dog

---

[1]One benefit of doing text trees is that it limits the extent to which various problem areas can be ignored. It limits the extent to which certain complexities can be pushed aside as belonging to some other system.

[2]Newari is a Tibeto-Burman language spoken as mother-tongue by half a million people who live in Kathmandu Valley and in major trading centers throughout Nepal. It has been heavily influenced by both Sanskrit and Maithili and has been in use as a literary language for several hundred years.

[3]Khicāyā macāta 'The children of the dog', appeared in the first volume of a three-volume collection of Kansakar's stories entitled Sasumāā and was published in Kathmandu by Himanchal Pustak in the year B.S. 2028 (A.D. 1971-1972).

This string serves as the title of the story. It is a term referring to a set of participants. One could presumably supply the predication of which it is a term (such as "This is a story about ___.") but we refrain from doing that at this stage.

To start with we assume that this is a noun phrase. If the tree assigned to this title recurrs frequently (with the same semantic correlates for its internal structure) as non-suspect noun phrases, our assumption will be confirmed. If not we may wish to explore some alternative analysis. Titles may rate special treatment.

A noun phrase may be either an item noun phrase with some kind of structural unit as head, or it may be an abstract noun phrase, perhaps derived from a clause, in which the structural unit functioning as head is either complex or missing. Thus in English we have both a) The book I borrowed yesterday was a good one. where book is a countable structural unit functioning as head and b) For me to borrow a book was probably a mistake. where the whole derived clause functions as a head and is not easily analyzed further in terms of relations normal to the noun phrase, but is better analyzed in terms of relations normal to the clause. The kinds of relations we normally find among constituents of a noun phrase include the head of the noun phrase, quantifier of the noun phrase, qualifier of the noun phrase (includes all non-quantitative modifiers, and may be broken down further), specifier of the noun phrase, and relational markers of the noun phrase.[4]

In the example above we select initially macāta 'child,Pl' as the head. The word, khicā-yā 'of the dog', is morphologically marked as a genitive and serves as a kind of specifier or identifier of the noun phrase. Specifically it marks kin relationship. In a wide range of languages (Indo-European,[5] Tibeto-Burman, and Malayo-Polynesian at least) the genitive or possessive form covers a fairly wide range of semantic relations. Because of this wide semantic range, some

[4]For a fuller summary of relations normally found within noun phrases see David Thomas, 1977, Noun phrase components, in R. Loving (Ed.) Proceedings of the S.I.L. Consultants Seminar, Ukarumpa, 1976 (Workpapers in PNG Languages, Vol. 20).

[5]Koiné Greek may be an extreme example. J. H. Greenlee, 1963, A concise exegetical grammar of New Testament Greek (Grand Rapids: Eerdmans) pp. 28-31 lists and exemplifies no less than 15 different uses for the genitive in Koiné.

some people are reluctant to use a term such as possessive because of its potentially diverse semantic implications. At this point one may elect to use 'genitive' as a more neutral designation. Hale at this point chose to use possessive in a grammatical sense, realizing that he might have to pick up some semantic pieces later on. The possessive form is itself fully expandable as a noun phrase. (What we know of the language as a result of language learning and as a result of preparing the accordian comes into play constantly in the process of treeing.)

The plural morpheme, however, applies to the head noun, and thus to the noun phrase as a whole. The head noun is unmarked for case (drawing again on what we know of the forms). Unmarked case is labelled as <u>Nom</u>, nominative. These considerations lead to the following surface tree.

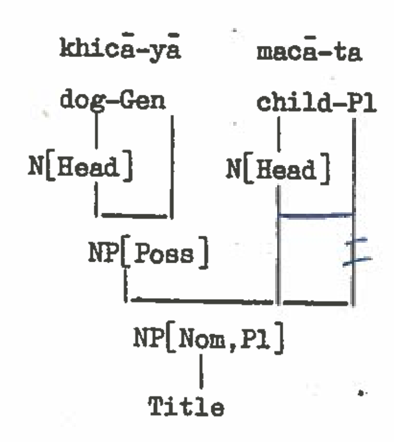


Figure 1.

The discourse constituent label, <u>Title</u>, need not wait for a later pass. It may, in fact, eventually end up as both the Cell 3 role of the noun phrase and as the name of the Cell 1 slot which it fills, in which case it would appear in the square brackets: NP[Title, Nom, Pl].

The label, NP[Poss] for <u>khicā-yā</u> is incomplete. The whole four-cell tagmeme could be represented somewhat as follows:

Poss	NP
Kin	[Gen]

but at this point Hale's laziness prevailed. We are not compelled to record all details at all times if they are not within the focus of our interest.

What we have done to this point is not simply assigned a tree to the first string of a story. We have the string to the term-type, NP, and we have started to specify the range of trees that can be NP's in Newari. If we encounter the tree-type exemplified in Figure 1 as a normal recurring structure in Newari texts it will tend to support this analysis. If this structure never shows up again, or does so only rarely, we may want to look for another tree structure that will give us a higher descriptive return for the space it takes in the grammar.

Consider now the first sentence of the text.

1. Cha-guu desa-e cha-mha mahārānii du.
   one-Cl country-Loc one-Cl queen is
   . . .
1. In a certain country there was a certain queen.

The terms in this sentence appear to be cha-guu desa-e 'in a certain country'; cha-mha mahārānii 'a certain queen'; and du 'there was'. The first two terms appear to be noun phrases; the last, a verb phrase (our work in constructing the accordian tips us off to the fact that this term is expandable as a string of words which fill the same predicate slot).

Consider first cha-guu desa-e 'in a certain country'. The whole term functions as a specification of location. (It may also do double-duty in marking the discourse as a folk-tale.) The locative marking is a clitic attached to the final content word of the phrase. It functions as a relational marker for the phrase as a whole. The first word, cha-guu is an indefinite quantifier consisting of a numeral followed by a numeral classifier. Comparing cha-guu with cha-mha in the following term we might guess that -guu goes with inanimate quantified heads and -mha goes with animate quantified heads. These considerations lead to the tree given in Figure 2.

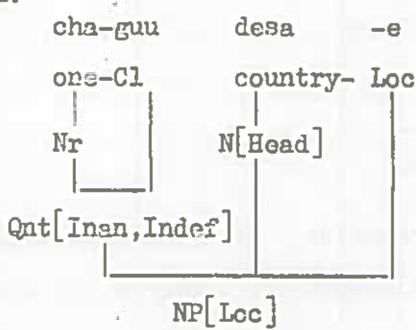


Figure 2

The label, NP[Loc] which has been assigned to this term is basically a Cell 2 label. The other cells are relational in nature and cannot be defined until we have looked at the next higher structure. If this is eventually interpreted as a peripheral item, the label, NP[Loc], would probably remain as is and our norms would state that this is a reduced form of

Mar	NP[Loc]
Loc	[Loc]

.

Otherwise a more complete set of features will be called for within the square brackets.

One may well ask why the locative, -e, is treed as a constituent of the noun phrase rather than as a constituent of the head word in Figure 2. The answer is that it could have been treated as an affix with some justification. The choice to make it a phrase constituent was based, however, on the fact that the case markers in Newari function as phrase clitics, rather than as noun head affixes. They may attach to any phrase-final content word. They are not grammatically bound to the head, but rather, to the phrase as a whole.

Consider now the second term of Sentence 1. The tree that we set up for the first term of Sentence 1 has immediate payoff for the second term. Except for the fact that the second term is unmarked for case rather than marked for locative, the tree is much the same.

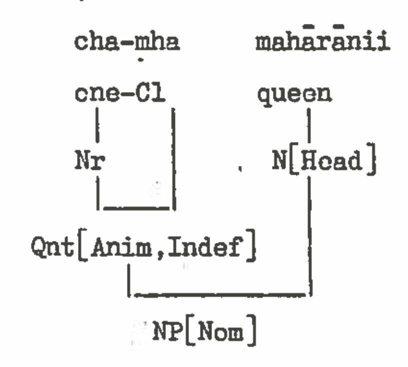


Figure 3

Again the label,NP[Nom], is incomplete, pending the results of our look at the clause.

The last term of Sentence 1 consists of one word, du 'there is', which is the existential verb. This verb is commonly used for bringing participants on

stage and for predicating existence, possession, or location of a subject item. In this particular instance [Existence] is the relevant feature.

du
is
|
V[State]
|
VP[Pred,Existence]

Figure 4

The morphological form of the verb is an irregular stative form, the infinitival form being da-ye. This information stems from morphological analysis that is best carried out on sets of paradigms and will not be discussed at this point.

Consider now the whole sentence. It consists of one predication or proposition, namely, that in a certain country a certain queen existed. Our initial assumption is that it consists of a single clause. In approaching the analysis of a clause we bring certain expectations to bear. Internal to the clause we expect to find a certain normal set of slot relations such as:

[Subj]	subject of clause	[Pred]	predicate of clause
[Obj]	object of clause	[Compl]	complement of clause
[Adjn]	adjunct of clause	[Mar]	margin of clause

We also expect to find a certain normal set of role relations such as:

[Act]	actor of clause	[Item]	item of clause
[Und]	undergoer of clause	[Stmt]	statement
[Site]	site of clause[6]	[Cmd]	command
[Inst]	instrument of clause	[Ques]	question
[Time]	time of clause	[Ben]	beneficiary
[Loc]	location of clause		

---

[6] Site here is equivalent to Scope in Pike and Pike, 1977, Grammatical analysis. The label, Site, betrays a slight localistic bias. See John M. Anderson, 1971, The grammar of case, towards a localistic theory (Cambridge studies in linguistics, Nr. 4) to gain an appreciation for what a localistic theory of case can offer. See also Joseph E. Grimes, 1975, The thread of discourse (Janua Linguarum, series minor, Nr. 207) The Hague: Mouton, Ch. 8, to get an idea as to how orientation and process interact within a case system.

Furthermore, we expect that the possible combinations of term-types within any given clause is controlled in part by the clause (or predicate) type. There are certain normal sets of predicate type which we expect to find in any language:

- Transitive set (including [BT] bitransitive; [T] transitive; [ST] semitransitive; [I] intransitive)
- Receptive set (including [BR] bireceptive; [R] receptive; [SR] semireceptive; [E] eventive)
- Stative set (including [BS] bistative; [S] stative; [SS] semistative; [D] descriptive)
- Attributive set (including [BA] biattributive; [A] attributive; [SA] semiattributive; [C] circumstantial)

As sub-types of the attributive set we have identificational and existential types as well.[7] We take the predicate to be the central, governing constituent of the clause, occupying a position within the clause analogous to that occupied by the head of the noun phrase.

Bringing these assumptions to bear on Sentence 1 in the light of prior work done on the language we find that the verb da-ye 'to exist, to be in a place, to have' is a biattributive existential verb and that the locative noun phrase it goes with is a nuclear adjunct-site and that the unmarked (nominative) noun phrase is a nuclear subject-item. The tree for Sentence 1 is as follows in Figure 5.

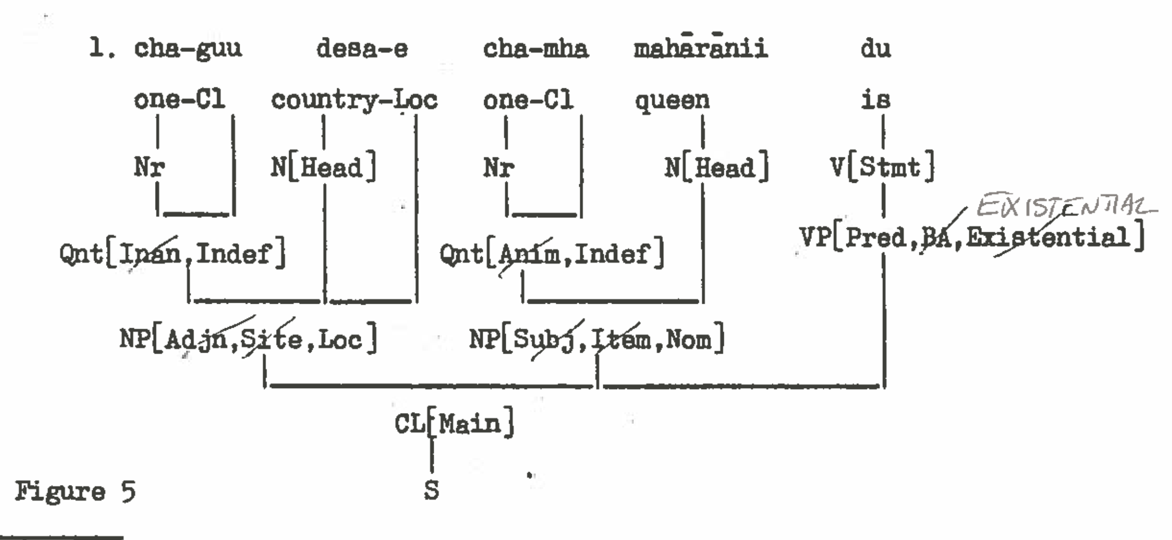


Figure 5

---

[7] For further discussion see A. Hale, 1974, On the systematization of box 4, in Brend (Ed.) Advances in tagmemics (North-Holland Linguistic Series, Nr. 9) Amsterdam: North-Holland Publ. Co., pp. 55-74; A. Hale, 1973, Toward the systematization of display grammar, in Hale (Ed.) Clause, sentence, and discourse patterns in selected languages of Nepal, (SILP, Nr. 40, Part I) pp. 1-38.

Figure 5 is rather cluttered. When this sentence was originally treed, certain norms were known that allowed nearly everything in square brackets to be omitted. Retained were [Indef], [Loc], [Nom], [Pred], and [Existential]. The rest was predictable. Tastes will differ as to how much information to display in the trees.

Consider now Sentence 2 of our text. Sentence 2 is potentially ambiguous,

2. wa maharānii-yā macā pwātha-e du-gu juyā cwana
that queen-Gen child womb-Loc be[L-gu] happen[PC] stay[PD]
. . .
2. As for that queen, it happened that she was pregnant.

the other interpretation being, 'That queen's child, as it happened, was in the womb.' On the first interpretation there are three noun phrases and a verb phrase.

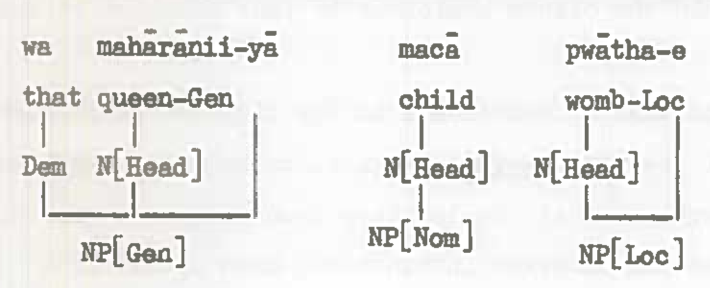


Figure 6.

On the second interpretation there are two noun phrases and a verb phrase.

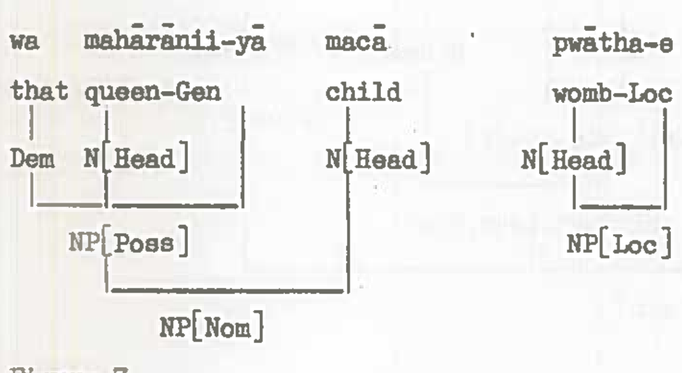


Figure 7

Both analyses are recorded on the accordian, though the first seems more likely due to the fact that <u>macā pwāthae daye</u> is an idiom 'to be pregnant' which happens to fit the existential-locative clause type in form. On this analysis, <u>wa</u>

maharanii-yā is a genitive marked sentence topic, 'as for that queen'. The auxiliary, juyā cwana has the force, 'it happened that' with overtones of 'lo and behold!' and is used a great deal by some story tellers and very little by others. The currently preferred tree for Sentence 2 is represented in Figure 8 in simplified form.

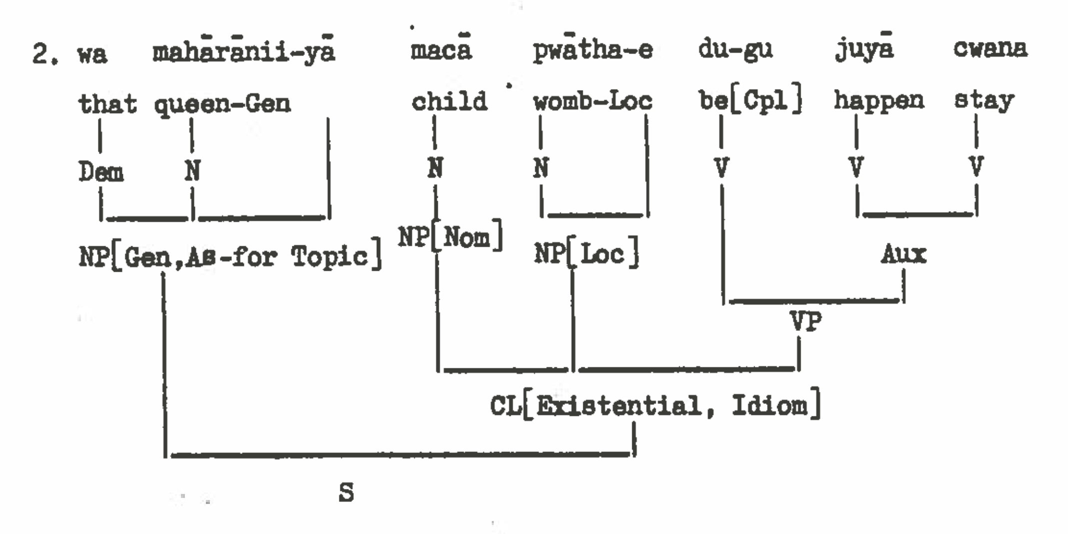


Figure 8. (Cpl = complementizer)

The tree in Figure 8 adds certain new elements to the set of structures posited for the title and for Sentence 1. For the first time we have a demonstrative within the noun phrase. For the first time we have a simple unmarked noun as a complete noun phrase. For the first time we have a genitive noun phrase functioning outside of the possessor slot. These kinds of things add up bit by bit to form a coherent picture of the noun phrase. At clause level we have a complementizer-auxiliary construction with a complex auxiliary. At sentence level we have our first as-for topic. Finally, we have seen an instance in which two analyses are at least technically possible and we have seen that the structural differences between the two correlate with semantic differences that can be tested.

Sentence 3 brings some more new items into view:

3. sarad ritu-yā yām
autumn season-Gen time
. . .

3. Autumn was the season.

In Newari, equative identificational clauses are typically verbless. Sentence 3 consists of two noun phrases, sarad 'autumn' and ritu-yā yām 'the season', the second of which functions as the identificational predicate. We posit the following tree.

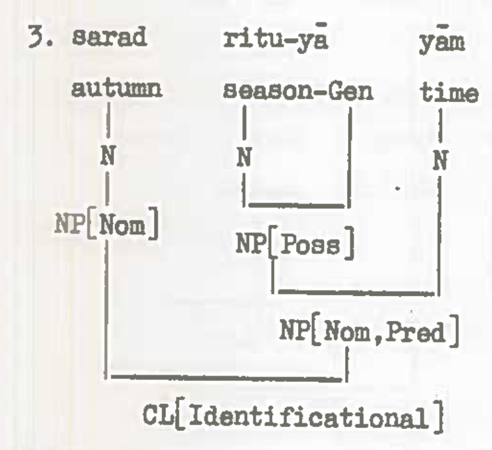
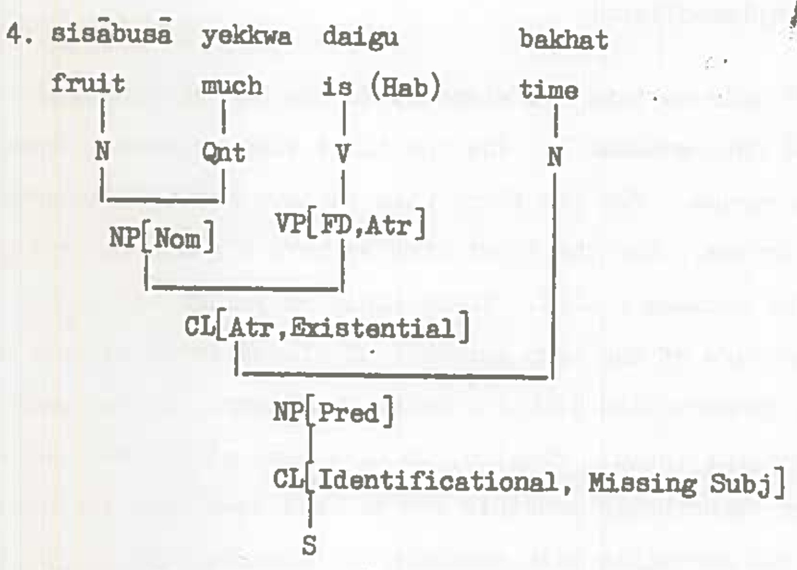


Figure 9

Sentence 4 appears to be an identificational equative clause with a zero subject.

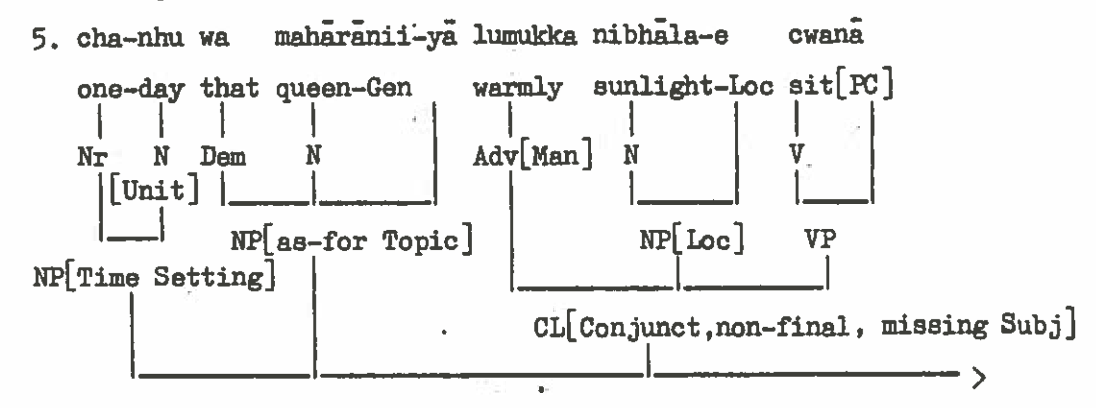


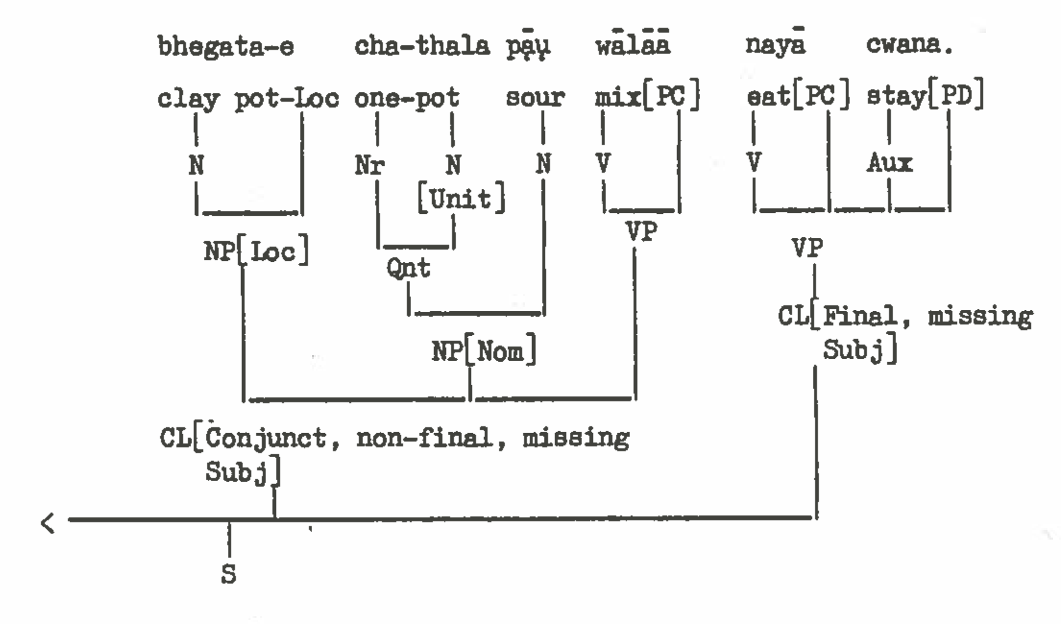
4. [It was] the time when there was much fruit.

Figure 10

Clauses with missing subjects are quite common in Newari texts. This sentence gives us our first example of a relative clause as noun phrase modifier.

Another very common type of sentence is exemplified by Sentence 5, which consists of a time setting followed by an as-for topic followed by a string of non-final conjunctive clauses and ending with a final disjunct clause.



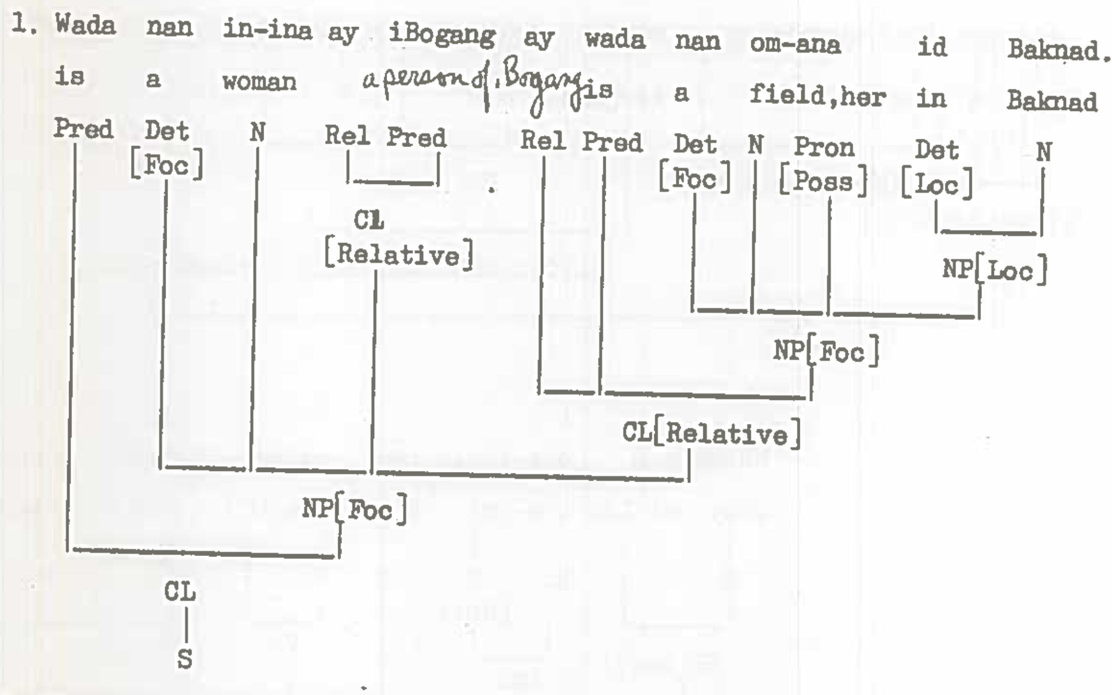


5. One day that queen sat warmly in the sun mixing sour [fruit] in a large pot-bellied clay pot and eating [it].

Once the setting has been presented, this general type of sentence becomes the dominant recurring pattern of the narrative. This little story in Newari continues for another 222 sentences. The text tree which has been constructed for it is in constant use whenever Hale works on Newari grammar. Before moving on to Step Three, however, it would be profitable to look at some text trees from a Philippine language. We start with the initial sentences of a text in Northern Kankanay as treed by Judy

Wallace. The text was told by C. Amkinit and is entitled, The time of the enemy.

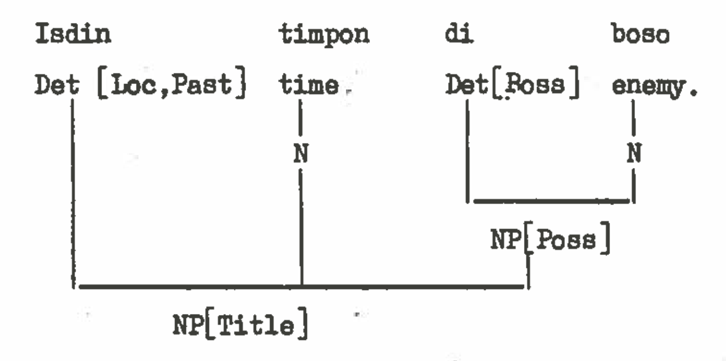
Sentence 1 is an existential, serving to introduce the main participant of the story. Two relative clauses are embedded within the focused noun phrase of the main clause.



1. There was a woman of Bogang who had a field at Baknad.

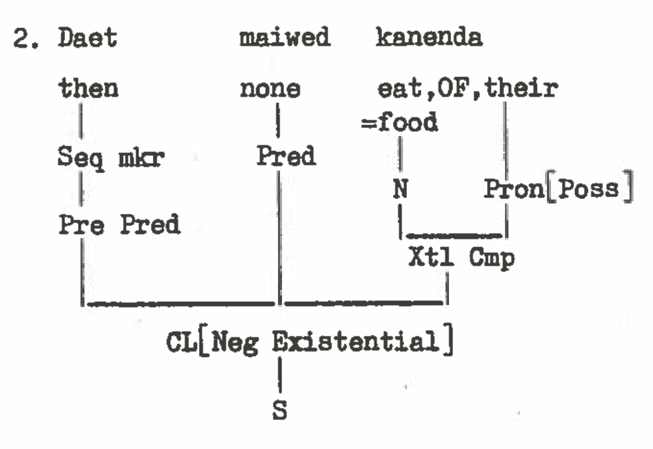
The relative clauses that are translated 'of Bogang' and 'who had a field at Baknad' both modify the noun phrase head, 'a woman'. The test for relative clause used in this analysis was the following: If the modifying clause could be converted into a main or independent clause by adding to it a noun phrase with a head identical to the modified head (<u>nan</u> <u>in-ina</u>) then the modifying clause is a relative clause. In other words, a relative clause has a missing noun phrase which can be supplied by the head noun of the noun phrase in which the relative clause occurs. This missing noun phrase is either the focused item of the clause or it is a preposed topic. The focused item is missing from the first of the relative clauses in Sentence 1. The preposed topic is missing from the second, and the possessive 'her' is in cross reference to this topic. Relative clauses with missing topics were found mainly with existentials, if Hale's memory serves him correctly.

It should be noted that the first relative clause is derived from an identificational equative clause with a predicate, iBogang 'a person of Bogang'. In this one sentence we find quite a few different relationships within the noun phrase. Besides the relative clause modifier we find both possessive and locative modifiers. The possessive exemplifed in Sentence 1 is a pronoun. In the title of the story we find a possessive with the form of a noun phrase.



The time of the enemy.

Sentence 2 is a negative existential. The internal structure of the complement of the existential (Xtl Cmp) is not altogether clear. It could prove to be a focused noun phrase with deleted determiner. Such deletions may prove to be common with negative existentials.

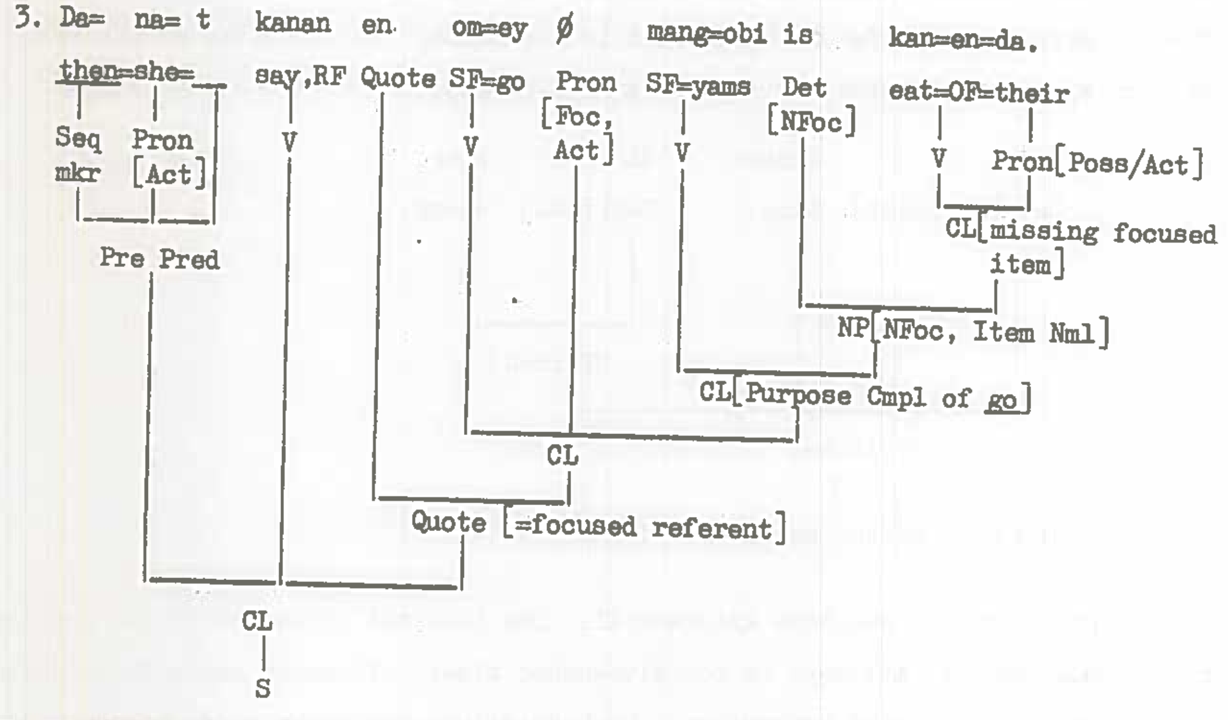


2. But then they had no food.

We encounter here for the first time the pre-predicate constituent, an important set of elements that mark sequence, negate predicates, and attract pronouns and particles to pre-predicate position.

Sentence 3 exemplifies not only a pre-predicate constituent with an attracted pronoun actor, it also exemplifies clause complement structures and it even has an item nominal derived from an underlying clause. (Equals sign signals morpheme break.)

3. Da= na= t kanan en om=ey Ø mang=obi is kan=en=da.



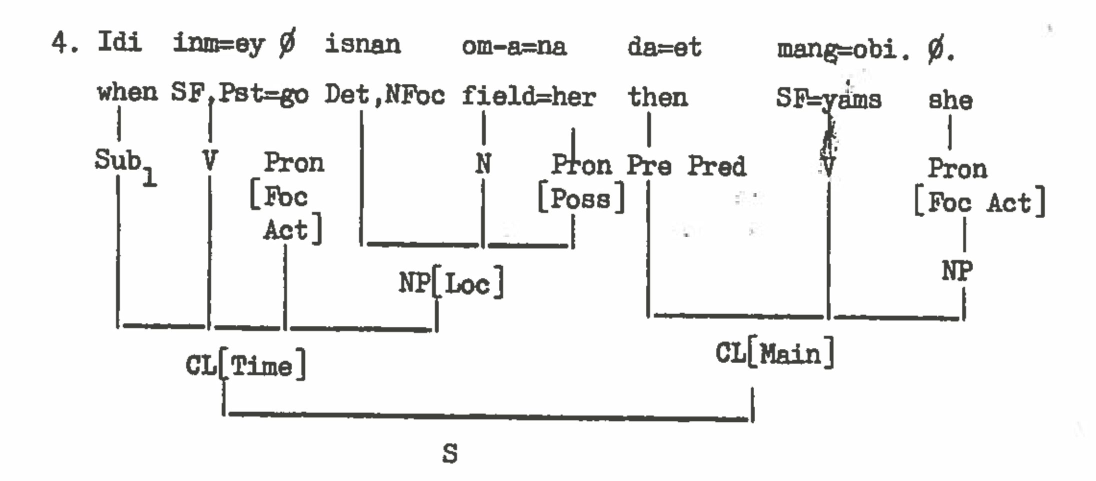
3. So she decided she would go gather yams for their food.

A number of things deserve comment here. The first word consists of daet 'sequence marker, so, then' plus an infixed pronoun actor, na, 'she' which was attracted from its normal position following the verb, kanan, 'said'. Notice the use of underlining to indicate the discontinuous morpheme, Da ... t in the gloss line.

The last noun phrase in this sentence is an item nominal, an extremely common type of nominal in the Philippine languages Hale has had a chance to look at. Item nominals consist typically of clauses from which the focused item is missing. The missing focused item serves as the semantic head of the noun phrase. In Northern Kankanay the determiner occurs with the item nominal construction.

Finally, the use of clause complements exemplified here is typical of the Philippine languages Hale has looked at. Two such complements occur here, one with kanan 'said', where the quote is a focussed item as well as a complement, and one with omey 'to go', where the complement also expresses purpose in a non-highlighted way.

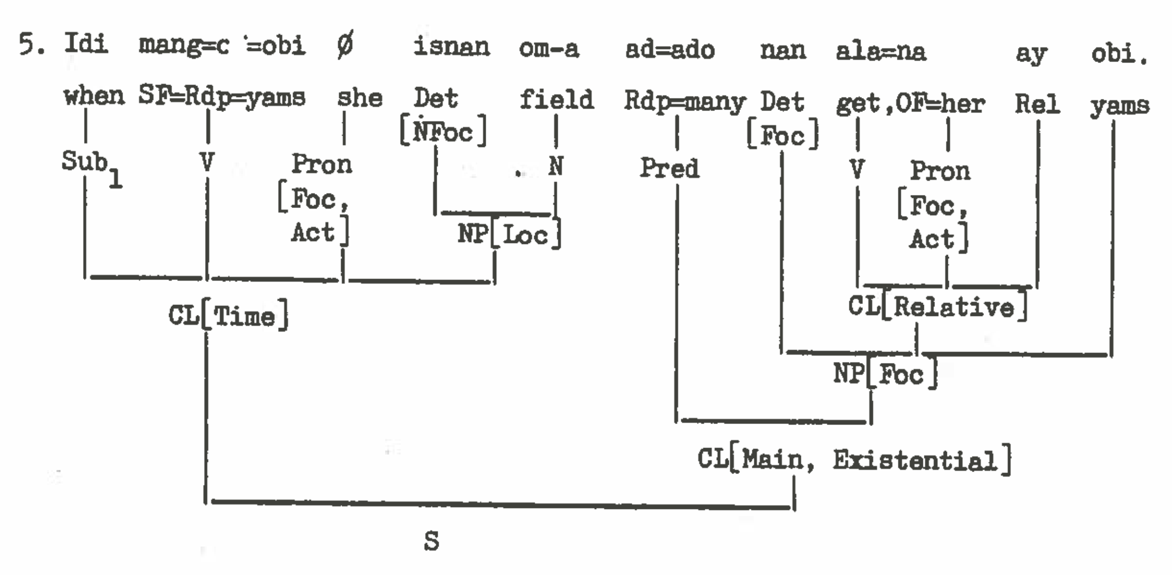
Sentence 4 brings us an example of a subordinate time clause.



4. When she had gone to her field, she gathered yams.

In Wallace's original draft the time clause was treated as a relater-axis construction. Hale is to blame for modifying the representation here, as well as in some earlier examples. The original can be made available for those who wish to see it.

It is not uncommon in Philippine languages that predicate quantifiers are used in an existential sense. Sentence 5 exemplifies such an existential.



5. While she was gathering in the field she got many yams (Lit: there were many yams which she got).

We shall have more to say about text trees in later lectures. This much should suffice, however, to illustrate the kinds of representation generally envisioned under the heading of text trees.

C. The Work Chart.

We consider now a kind of structural concordance to be derived from the text trees. The work chart begins with the text trees and moves toward a formal summary of the basic patterns of which the trees are a manifestation.

In constructing individual trees we tried to keep in mind the total system. Each tree either follows from the system or adds to it or points toward a reanalysis and revision of the system, or is simply an inconsistent alternative to be revised in the light of the system. In this step we tighten the relationship between our trees and the system.

The work charts envisioned here are laid out in terms of functional positions. The positions are labelled across the top of the chart. The occurring patterns are laid out in the columns. The functional positions are arranged so that the elements of the unit under study can be laid out in the same left to right order as they occur in the text. This is done even if it becomes necessary to lay out two or more columns in the chart with the same heading. In other words, alternative orderings of functional positions are handled here by repetition of columns. Often it will be possible to distinguish preposed and postposed positions outside the nucleus. Nuclear elements are often fronted or postposed for prominence. If prominence itself constitutes a functional position it is sometimes useful to set up such a column.

The following is an example of the workchart. It deals with the Newari noun phrase and consists of examples taken from the story, The children of the dog. In this instance it worked out quite well to have a single work chart for all noun phrases. It may well be the case in other languages, or for other constructions in Newari, that more than one chart will be required to arrive at a useful result. We have found it useful to go through a text sentence by sentence when making such a chart. In this way noun phrase variations can be correlated with discourse.

We shall have more to say about text trees in later lectures. This much should suffice, however, to illustrate the kinds of representation generally envisioned under the heading of text trees.

C. The Work Chart.

We consider now a kind of structural concordance to be derived from the text trees. The work chart begins with the text trees and moves toward a formal summary of the basic patterns of which the trees are a manifestation.

In constructing individual trees we tried to keep in mind the total system. Each tree either follows from the system or adds to it or points toward a reanalysis and revision of the system, or is simply an inconsistent alternative to be revised in the light of the system. In this step we tighten the relationship between our trees and the system.

The work charts envisioned here are laid out in terms of functional positions. The positions are labelled across the top of the chart. The occurring patterns are laid out in the columns. The functional positions are arranged so that the elements of the unit under study can be laid out in the same left to right order as they occur in the text. This is done even if it becomes necessary to lay out two or more columns in the chart with the same heading. In other words, alternative orderings of functional positions are handled here by repetition of columns. Often it will be possible to distinguish preposed and postposed positions outside the nucleus. Nuclear elements are often fronted or postposed for prominence. If prominence itself constitutes a functional position it is sometimes useful to set up such a column.

The following is an example of the workchart. It deals with the Newari noun phrase and consists of examples taken from the story, The children of the dog. In this instance it worked out quite well to have a single work chart for all noun phrases. It may well be the case in other languages, or for other constructions in Newari, that more than one chart will be required to arrive at a useful result. We have found it useful to go through a text sentence by sentence when making such a chart. In this way noun phrase variations can be correlated with discourse.

Newari Noun Phrase Work Chart

	Determiner/ Possessive	Quantifier	Qualifier/ Attributive	Item	Plural	Quantifier	Case	Emph
1.	khicāyā			macā	ta		Ø	
		chaguu		desa			e	
2.	wa	chamha		mahārānii			Ø	
				mahārānii			yā	
				macā			Ø	
3.				pwātha			e	
	rituyā			sarad			Ø	
4.				yām			Ø	
				sisābusā		yekkwa	Ø	
5.			sisābusā yekkwa daigu	bakhat			Ø	
	wa	chanhu		Ø			Ø	
				mahārānii			yā	
				nibhālā			e	
				bhegata			e	
6.		chathala		pāų			Ø	
				pwātha			e	
7.	wa		pwāthae du mha	khicā		chamha	Ø	nąą
				khicā			ą	
				mahārānii			yā	
				pāų			Ø	
			mahārāniiyā pāų walāā					
			nayā cwąągu	Ø			Ø	
8.				i			Ø	
				mahārānii			yā	
				me			Ø	
				pāų			Ø	
	wa			khicā			yā	
							yātā	nąą

This kind of work chart is a useful display to consult for answers to a surprisingly large range of questions, not only concerning the internal structure of the noun phrase, but also concerning the effect of discourse upon the variant forms noun phrases take throughout a discourse.

This fragment of the Newari noun phrase work chart raises as well as answers certain questions. From this fragment it is possible to see that determiners, quantifier, qualifiers and items all function as heads for the noun phrase. In general the last constituent before the plural slot is head but the last constituent before the case slot is inflected for case. This supports the view that case in Newari functions as a phrase level clitic.

Unanswered questions raised by the chart include the question as to whether plural can cooccur with the quantifier or not. Native speakers disagree on this. Can a qualifier head be quantified? Can a determiner head be quantified?

Not included in this chart are the zero references to participants that can be found in or deduced from the text. These constitute an important aspect of participant tracing patterns in Newari narrative so not all our discourse needs are served by this kind of chart. Pike and Pike, 1977, Grammatical analysis makes liberal use of work charts of this sort. The reader stands to profit from studying them.

D. Broader Coverage: Total Filing, Concordance Search.

Treeing texts and constructing workcharts is very useful as a starting point for identifying construction types at all levels of discourse. It is essential to be able to see a discourse as a whole and to analyze each part in relation to the whole. Only in this way can a modular approach be guaranteed of some degree of stability early in the analytic process. Treed text is a very useful kind of display both for low level structural analysis and for the study of whole discourses. One must be strategically selective, however, since to tree all of ones corpus of text could be prohibitively time consuming.

One needs to have broader coverage of a language than is offered by the relatively few texts one decides to tree. How does one get this kind of coverage more quickly than by treeing, yet in a form amenable to solid analysis?

Those who have computer concordances already have access to this kind of coverage. In working with Wallace and others Hale has observed that time spent with concordances paid off handsomely both in raising relevant issues early and in answering questions for which broad coverage of text was required.

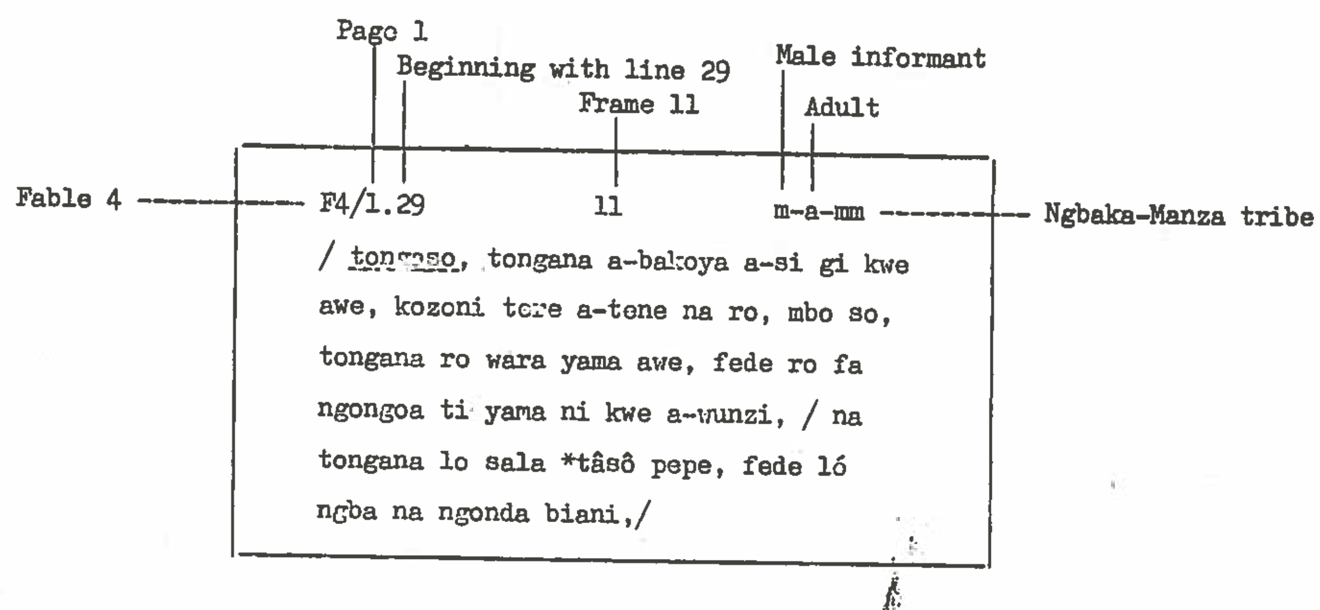
With the era of cheep concordances gone and the era of field-produced concordances yet a thing of the future for teams in the Philippines, for Step Four we turn to the possibilities offered by complete filing, a technique described in 1967 by William J. Samarin in his book, Field Linguistics, a guide to linguistic field work (New York: Holt, Rinehart and Winston) pp. 159-162. Samarin describes the technique as follows:

"The complete filing system is free of the disadvantages inherent in the selective system and enjoys other advantages. It is simply this: one takes a particular corpus and files away every conceivable bit of information. While it can be used for phonological as well as grammatical analysis, it is the latter use that we shall describe here. It should be remembered that complete filing can also be done with manual punch cards or electronic computers, but we are using the term exclusively for slip filing. The complete coverage of the corpus is made possible by duplicating machines, for it is the making of slips that largely determines how much filing is done. Once the slips are made, the collation proceeds as with any slip-using technique. In describing the process we must therefore pay greatest attention to the preparation of slips.

"One begins with the preparation of the master. Any duplicating process is adequate: spirit-duplication, mimeographing, offset. The first process is the least expensive and the simplest to operate. It can also be used for original transcriptions in the work sessions, in this way one can do all his work directly on the masters. Its disadvantage is that it produces a restricted number of legible copies. When working with texts, it is preferable to use the mimeographing machine: time and labor are saved by the use of the longer stencils; stencils can also be re-used if more slips are required. When properly cared for (between absorbent sheets or water-washed--if the proper kind of ink is used), stencils can easily be used after several months even for long runs. They should be clearly labeled. Incidentally, proof-reading of the stencil can be easy if one uses carbon cushon sheets which are sold by the larger distributors.

"The master is divided into 'frames' of whatever size is most convenient for storage and handling. With an 8 1/2-x-11-inch stencil one obtains 8 frames 2 3/4 x 4 1/4 inches in size. With a 14 inch stencil there are 10 frames of the same size, leaving a 1/4 inch strip of waste. Onto these frames sections from the text are typed in such a way that each frame has material of more or less the same length, and no sentences are avoidably left incomplete. All frames are numbered serially for each text, and the source of the data is clearly indicated. If someone other than the investigator is going to complete the preparation of the slips for filing, the material has to be coded for filing. For those working on the Sango grammar project, for which the following slip was prepared, the instructions were: underline

everything separated by a hyphen or word division; underline words marked with an asterisk twice (that is, make two different slips for them). The drawing below illustrates a single frame from which a slip has been prepared for the word tongaso. (The orthography in this illustration is different from the one later adopted and which is used for all citations in Sango in this book.) The next slip prepared from this frame was for tongana, the next for a-, and so on, until all words had been underlined, one to a slip (and *tâsô was marked twice), making a total of 45 slips. Then a duplicate set--another 44 slips--was made for the dictionary project. The slips which were left over--and extra ones were deliberately made by running off the stencil more than 89 times--were stored away for other uses.



"The hyphenated syllables were prefixes (Plural marker in a-bakoya and subject marker in a-si) and the words with preposed asterisks were French words of which a special study was going to be made. To reduce the number of slips we decided to write some words 'solid.' Although so, ni, and na are separable words, they were written solid in tongaso, kozoni, and tongana, because we knew that these forms were extremely common. If we wanted to recover those occurrences of so, ni, and na, we could go to the files for tongaso, and so on. The commas and diagonal slashes indicated short and longer pauses respectively. They were not used in the processing of data but were necessary for the syntactic analysis. This was principally a word and morpheme file. We could have coded the text for syntactical analysis easily enough, but this was not necessary. All noun phrases were recoverable from Nouns; verb phrases (as predicate, complements of other verbs) from Verbs; dependent clauses from tongana. For example, we had all the data we needed under Verbs to study the structure of two-verb constructions which are so common throughout much of Africa: namely, fede ro fa ngongoa ti yama ni kwe a-wunzi 'he will cut off (and) destroy (the) seed (descendants) of the animal completely.'

"Underlining has been suggested as the means to code the slips, because it is the easiest operation to perform: one simply fans out the handful of slips and then makes a line under each bit of information. This coding will be done neatly and unambiguously if there is sufficient room between the segments and between the lines

of material on each frame. This is why it is advisable to segment at the time of preparing the frames. Notice how much clearer is the coding of the Kikuyu verbal root [mberer] in A than in B; the latter represents segmentation done after the frames have been prepared:

A	B
ma-ge-ke-a-mberer-ia	ma|ge|ke|a|mberer|ia

It is obvious that the slip markers will have to exert more care, and therefore take more time, in markings the second utterance.

"To determine how many slips are needed for each stencil one simply tabulates the total of every item which is coded (that is, underlined) for filing. For this purpose it is useful to have stencil record sheets of the type which is illustrated below.

Stencil No. 1

Code No. L1/1

Frame No.	Words	Morphemes	French	Sentences	Total
1	26	0	1	2	29
2	28	2	2	2	34
3	33	1	1	2	37
4	29	1	1	2	33
5	23	6	2	2	33
6	22	4	0	2	28
7	24	0	3	3	30
8	24	2	0	2	28
9	19	2	0	2	23
10	19	1	4	2	26
Total	247	19	14	21	301

It is wise to have some extra slips for contingencies, such as inadequate duplication and unforeseen filing needs.. It is obvious from the figures above that frame 3 will require the production of more slips than is necessary for the other frames (14 more than frame 9 requires). This is the reason for attempting to make the contents of each frame as much alike as possible. But the extra slips do not constitute a great waste. Each slip costs only a fraction of a cent, an insignificant factor when compared with the efficiency of the technique. (Incidentally, the stock of unused slips should be well labeled and stored away near at hand. In the United States large quantities of shoe boxes of the exact width of the slips were easily obtained for this purpose. Specially cut but unmounted cartons can also be ordered commercially for this purpose; they can be sent flat to the field.) One colored sheet should be run off to be cut up into dividers for separating each set of frames after they are cut. There should be a few sheets left uncut. These are useful in the study of prosodic and other features where a connected text is needed. After the slips have been cut, being careful to make them of uniform size, they are ready for underlining. It is this underlining which identifies the piece of information which must be filed. Unskilled labor can perform this task. Among the more than 74,000 slips underlined for the Sango project by a group of housewives, duplications and omissions were rare indeed.

"When many different texts are being processed in this way, it is wise to keep complete and up-to-date records of the progress being made on each one, especially when there are assistants who are responsible for some of the work. But even when one is working alone, it is easy to lose track of what one has been doing. A progress chart of the kind shown below is strongly recommended. In each cell one adds the date of completion.

	F1	F2	F3	F4
1. Transcribed				
2. Typed				
3. Marked for stencil				
4. Stencil prepared				
5. Text run off				
6. Slips cut up				
7. Slips underlined				
8. Slips alphabetized				
9. Vocabulary filed				
10. Grammar filed				
11.				

. . .

"Among the many advangates of complete filing the following can be mentioned:

"(a) It can be initiated at any stage of the field work. It is as useful in working with material still poorly transcribed and inadequately analyzed as it is with material at the later stages of analysis.

"(b) It can be done under rather primitive conditions with untrained help.

"(c) It can be used for several different projects at once (dictionary and concordance filing as well as phonology and grammar).

"(d) It provides the analyst with a large portion of linguistic context for each bit of information.

"(e) It is economical with human labor and in terms of the equipment and materials used. In processing anything up to around 50,000 words it therefore has much more in its favor than the next technique, edge card. For a larger corpus electronic

computers are probably advisable.

"(f) It can be used with a small corpus as easily as with a large one. For example, a Temne fable of only 500 words was filed in this way in teaching linguistic analysis to a student. It is in fact advisable to process some data experimentally at first. This analytical experience might reveal the need for coding the texts in more, or less, elaborate ways." [Samarin, 1967, pp. 159-163]

Ken Maryott has had a great deal of profitable experience with the technique of complete filing and he will return to this subject in a later lecture.

E. <u>The Formal Summary</u>.

Each horizontal line in the text tree corresponds to some construction type (alias, syntagmeme, constructional root / stem) at some functional threshold. Step Five is concerned with providing a workable summary of the trees for each construction type at each functional threshold.

Any theory that provides a way to summarize the possible variants in a surface structure string can be used as the basis for representing the formal summary. Tagmemic formulae of the four-cell variety are perfectly suitable for this purpose. This kind of formal summary is explained and exemplified in detail in Pike and Pike, 1977, <u>Grammatical analysis</u>.

Another kind of formal summary that is less well known and is not nearly as easily understood from the literature is that of network grammar. Network grammar is not a linguistic theory. Rather, it is an alternative way of writing a grammar for any theory that incorporates a representation of surface structure. One major source of information on this is Joe Grimes, 1975, <u>Network Grammars</u> (SIL Publications in Linguistics and Related Fields, Nr. 45) Norman Oklahoma: SIL. Many have found this difficult to read. For those who wish to work their way through Chapter 3, 'Transition network grammars: a guide', we are including a glossary which Larry Seaward compiled while working to understand that chapter.

Network grammars appeal strongly to some people. Whether one finds this kind of representation congenial or not seems to depend on rather personal reactions. One need feel no guilt if one does not find this mode of description exciting. One must, however, find some means for summarizing the surface patterns of ones

language in a consistent and testable manner. Barbara Friberg has done very useful work on Cham with this mode of description.[8] Ross Errington has summarized a great deal of what has been done on Cotabato Manobo within this kind of formal framework.[9]

One of the attractions of this kind of grammar lies in the fact that the abstract underlying structures required by any theory can in principle be built by such a grammar and that such a grammar can be written in a computer programming language called LISP and verified by computer against text. This kind of attraction is stronger, of course, where field workers have access to large computers than where they do not.

The variant of network grammar introduced briefly here is something less than a full-blown LISP program. It is intended primarily as a compact visual display of surface structure patterns at all levels from phrase to monologue. The emphasis here is upon the utility of networks as an aid in organizing field research. One consequence of this emphasis is that several kinds of informal abbreviational conventions are presented that would look rather different in a more formal representation.

Each line of the Newari Noun Phrase Work Chart (Page 37, above) corresponds to one horizontal line in the text tree (though not all lines are actually represented in that version of the work chart). Each column corresponds to one functional position on the horizontal line. Consider again the tree for the title noun phrase together with its full workchart representation and its network summary as presented in Figure 11. Notice that the tree actually contains two noun phrases, one embedded as the possessor of the other. Thus the tree contains two horizontal lines, each of which corresponds to a noun phrase. Accordingly there are two lines in the work chart, one for the possessive noun phrase and another for the nominative noun phrase in which it is embedded.

How does the network manage to summarize this kind of embedding? It does so in a way that is quite important to understand. One pass through the network

---

[8]Barbara Friberg, 1978, Augmented transition network of the Cham language (University of Minnesota Thesis for the M.S. in computer science).

[9]Ross Errington, in press, A transition network grammar of Cotabato Manobo, to appear in Studies in Philippine Linguistics 3.2.

is basically a summary of the possibilities for any one horizontal line in the text tree. The noun phrase we are looking at in Figure 11 requires two passes through the network. Just as one noun phrase is embedded within the other, so

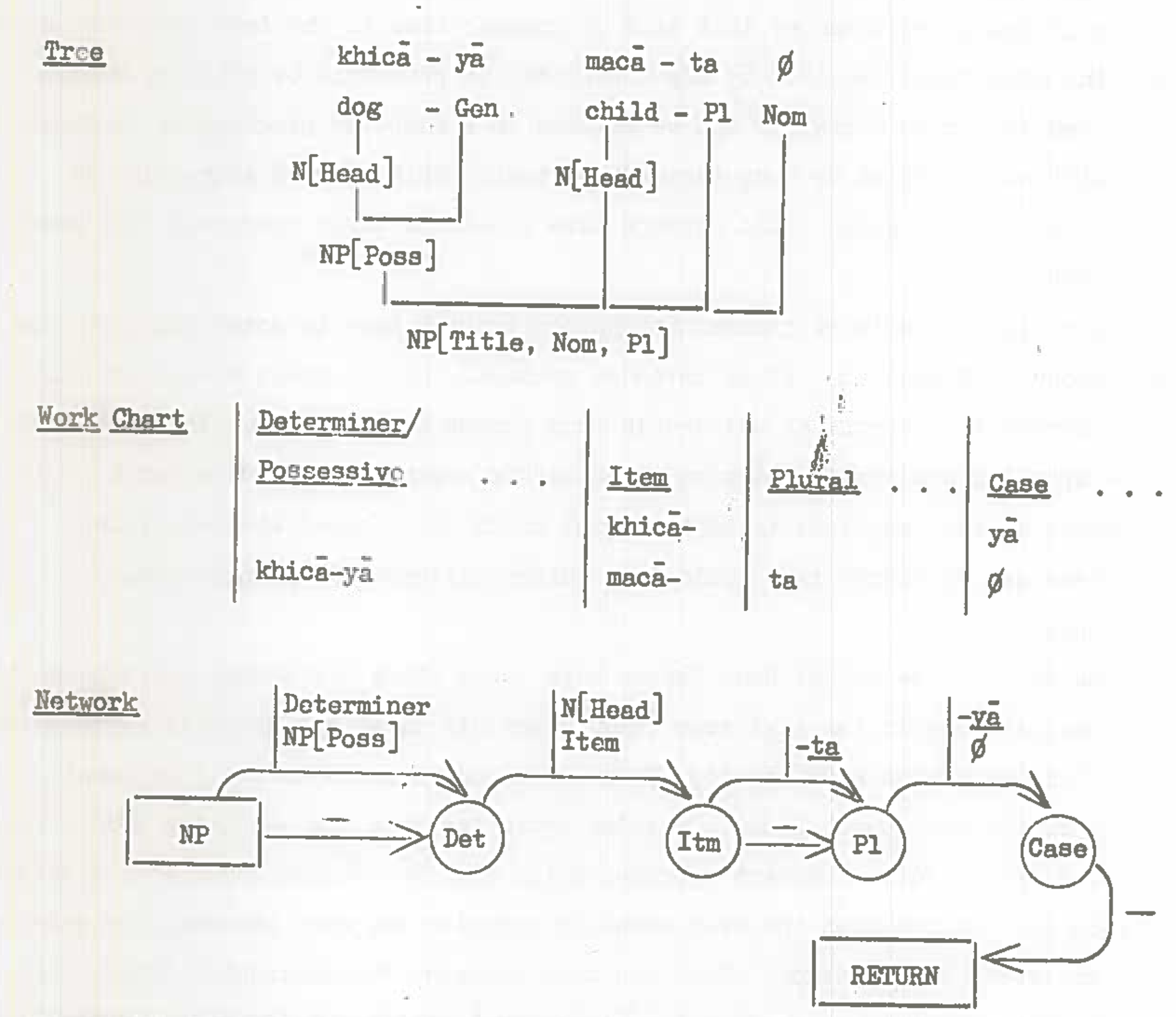


Figure 11. The relationship between text tree, workchart, and network.

also one pass through the network is embedded within another. The network can be conceived of as a parser that reads the text and assigns a tree to it. While reading the text the network makes use of a dictionary that assigns glosses and part-of-speech designations to the morphemes as they are read. Figure 11 gives only that part of the total network needed to read noun phrases of the kind given in the title for which the tree is given. NP is analogous to the beginning of a horizontal line so labeled in a tree and the box RETURN is analogous to

the end of the horizontal line. It is an instruction to return to the next lower line of the tree and to continue reading along that line by matching constituents of the text with the labels on the arrows (technically 'arcs') in the network.

The network in Figure 11 treats the categories, Determiner, NP[Poss], and -ta as optional since it is possible to bypass these constituents on arrows ('arcs') marked with a minus. We shall speak of such arrows as free passes. Where there are no free passes the constituent is treated as obligatory. Arrows will be referred to as arcs and circles will be referred to as states. Labels for states are normally either slots or roles. Labels for arcs are normally classes from Cell 2.

The labeled boxes that are analogous to the beginning of a horizontal line in a tree (and from which the arrows start) are called initial states. Thus NP in Figure 11 is an initial state. This initial state can be reached (or, more technically, 'called') from any label, NP, that appears on an arc. This, then, is an important way of embedding one pass through the network within another pass. Consider how this works on the tree of Figure 11. We start reading at the beginning of the line labeled NP[Title, Nom, Pl]. The first constituent we read is labeled NP[Poss]. This entitles us to move from the initial state, NP, along the arrow labeled NP[Poss]. Before we can move all the way to the state, (Det), however, we must verify the match between the label on the arrow and the structure in the tree. To do this, we hold our place on the lower line and move to the next higher horizontal line of the tree and start reading that line (matching it against the arc labels of the network) by entering the network again at state NP. The matching of the structure of NP[Poss] with the network succeeds by matching khicā with N[Head] and yā with -yā (which is the only match allowed here for a NP[Poss]) and by taking free passes for the rest. Once the higher horizontal line has been successfully read by the network we return to the lower line and resume reading at (Det) since we have already parsed the first constituent on the line, NP[Poss]. The reading of the lower line matches macā with N[Head] on the ard leading from (Det) to (Itm). It matches -ta with the label on the arc from (Itm) to (Pl). It matches Ø with the Ø on the arc from (Pl) to (Case). With this matching complete we can say that the network accepts the noun phrase as well formed and that it includes its structure.

2. Development of an Outline

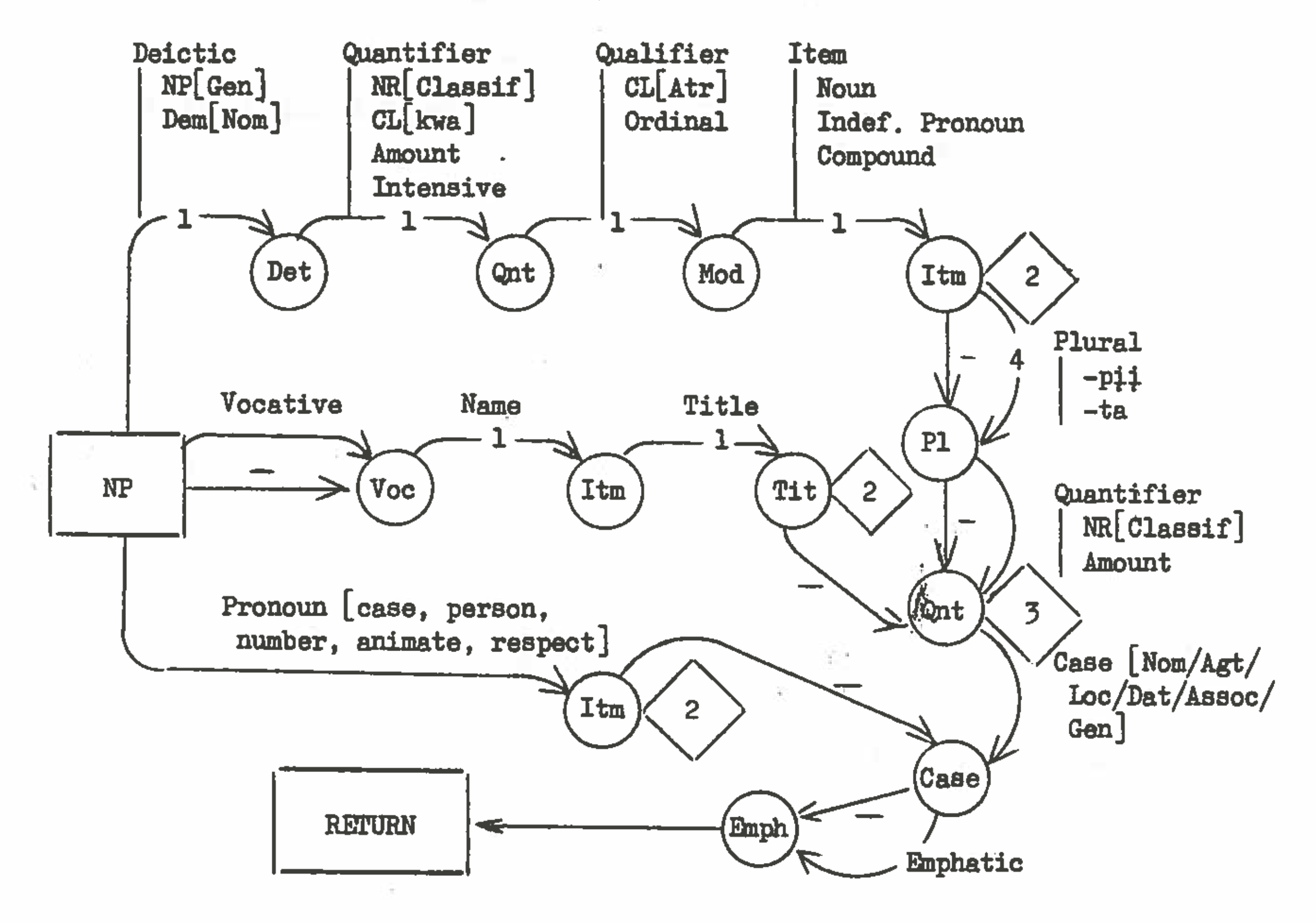
Properly used, networks are capable of a very direct, testable and compact summary of text trees. Austin Hale has had a good deal of happy experience using networks as a basis for constructing the outline of a thumbnail sketch. He will deal with networks in more detail in a later lecture. The reader will have gained enough of an understanding of what a network is and how it can be used to make it possible to build on the idea of the network in the following section.

F. The Outline.

The final step we wish to take in this lecture brings us from the formal summary to the outline of the thumbnail. The ideal outline for a loose leaf grammar notebook is derived not from some etic framework or abstract theory but rather from the structure of the language itself. The ideal outline is modular, being divided into various functional thresholds or levels and on each such level it is further divided into construction types. Added stability for the outline is built in by basing the outline on surface forms rather than on abstract underlying representations that are liable to change radically as the analysis progresses. We have suggested further that the outline be tied to the grammatical hierarchy and that the semantics or referential hierarchy be in large measure approached or constructed in the form of a topical index for material in the grammatical hierarchy. The independent substance of the index will, of course grow as the sketch matures, and if the index is well designed to begin with it will grow into what may eventually look more like an independent section heavily cross referenced to the grammatical section.

How, then, does one derive an outline for a thumbnail sketch from the surface structure of a text? The answer we propose here is that the outline comes from the formal summary. The particular kind of formal summary used for illustrating this answer is network grammar, but sets of related four-cell tagmemic formulae could probably serve equally well. The illustration is taken from Hale's thumbnail sketch of Newari and this one was chosen simply because it was closest at hand at the time of writing. We start with the assumption that any thumbnail will have a major section for the noun phrase. The outline was carried down to this point

in the first lecture. The outline for the noun phrase section is derived from the noun phrase network. Figure 12 is the opening page of the noun phrase section of Hale's Newari thumbnail. On this page is the noun phrase network, footnotes to the network, and the outline of the noun phrase section as derived from the network.



Section Outline

Noun Phrase General	NP0
Deictic	NP1
Quantifier	NP2
Qualifier	NP3
Item	NP4
Vocative	NP5
Name and Title	NP6
Number and Case	NP7
Pronoun	NP8
Emphatic	NP9

Notes:

1. Constituent is optional but at least one 1-marked constituent must be chosen within the chain.
2. The last constituent chosen by this point is HEAD.
3. Quantifier can be chosen only once within a given pass through NP. (Post-head Quantifier possible only if Item is head (?))
4. If the head is a Quantifier, Plural cannot be chosen.

Figure 12. Opening page of the Newari Thumbnail Noun Phrase section.

Pagination is by subsection rather than consecutive for the section as a whole. Thus Noun Phrase General starts on page NPO.1 and continues in the current draft through NPO.5 but additional pages could be added after NPO.5 without requiring any repagination at all. One should also feel quite free to add 'a' and 'b' pages (such as NPO.1a, NPO.1b) as needed. Deictics start on page NP1.1 and continue through NP1.5. In this series of pages NP1.3 and NP1.4 have been left blank for the time being to allow some additional space for the discussion of demonstratives which starts on NP1.1 and NP1.2. Either of these sections can be extended indefinitely without repaginating adjacent sections. As the sketch now stands the Quantifier section runs from NP2.1 through NP2.42 with certain built-in gaps, but this is the most extensive of the subsections as the thumbnail now stands. Section NP3, which deals with Qualifiers promises to be a big one since it will contain the discussion of relative clauses in their roles as modifiers and heads of noun phrases but section has not yet been developed very far.

We shall have much more to say about the development of a thumbnail sketch in subsequent lectures. For this lecture we will limit ourselves to illustrating two crucial recommendations for the outline and arrangement of the thumbnail. The first of these is the recommendation that summary networks such as the one given in Figure 12 always be followed immediately within the section with enough illustrative material and discussion to make the description useful and clear even to someone who cannot understand or refuses to look at network diagrams (this will help even the diagram lover from forgetting what his beautiful network stands for and just why it was laid out the way it was). The second recommendation is that subsections which have anything more than single-word structures to deal with also be introduced in the same way as the noun phrase section illustrated in Figure 12--with a network, footnotes, and a sub-section outline, the first sub-sub-section of which is general and illustrates in full the range of structures covered by the introductory network.

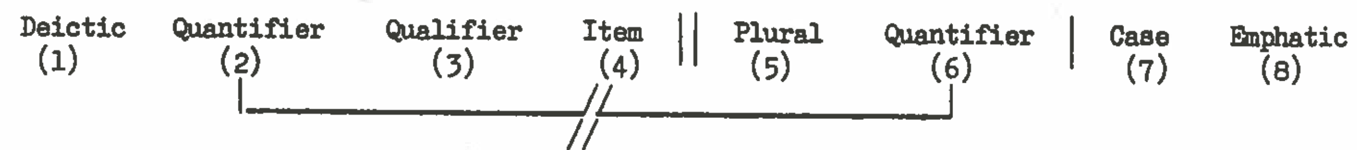
As an illustration of one instance in which the first recommendation was attempted, we reproduce several pages in the Noun Phrase General section of the Newari thumbnail sketch which immediately follow the page reproduced in Figure 12. Bear in mind that these pages were written not for publication but solely for the benefit of the writer and his partner in gaining an explicit understanding of the structure of the noun phrase in Newari.

Noun Phrase -- General NPO.1

"There are three main paths through the noun phrase network. Each path corresponds to a major NP construction type. We will refer to these as 1) General NP (corresponding to the upper path), 2) Proper NP (corresponding to the middle path), and 3) Pronominal NP (corresponding to the lower path).

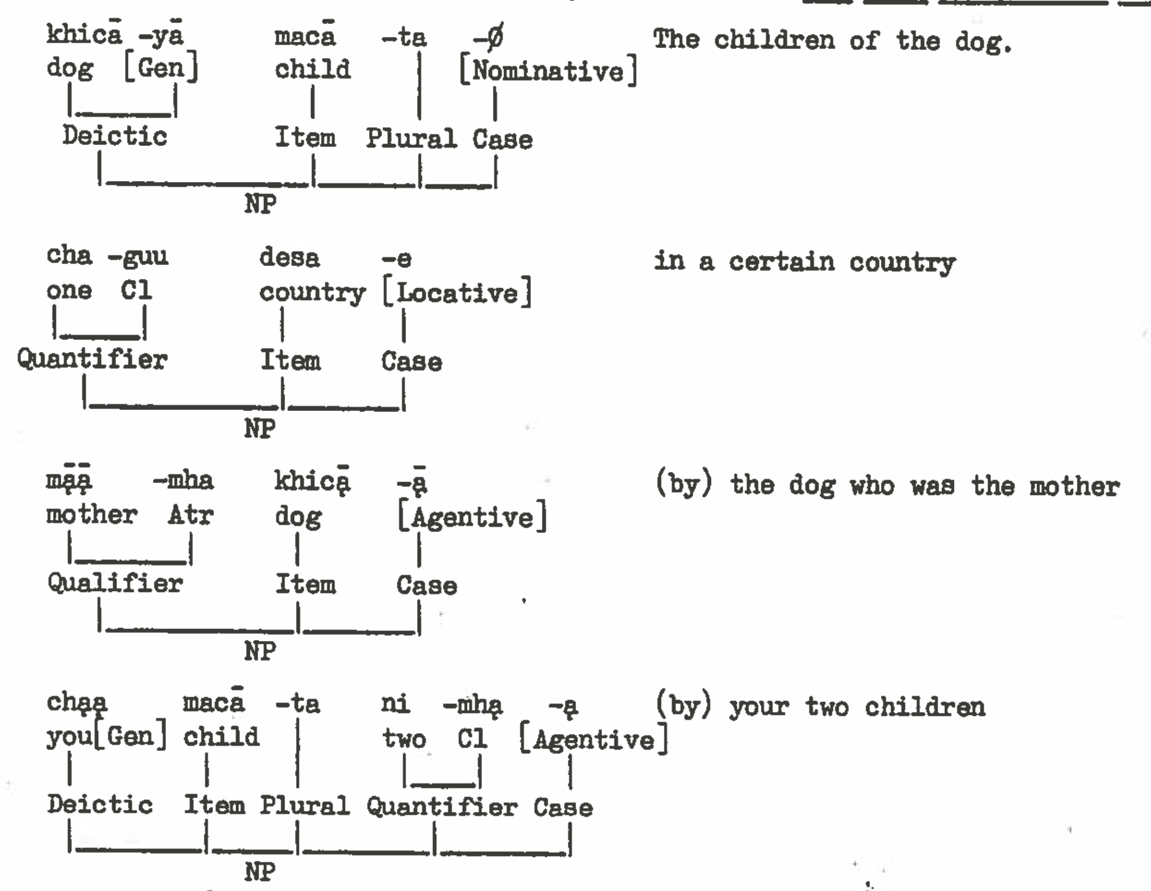
A. The General NP

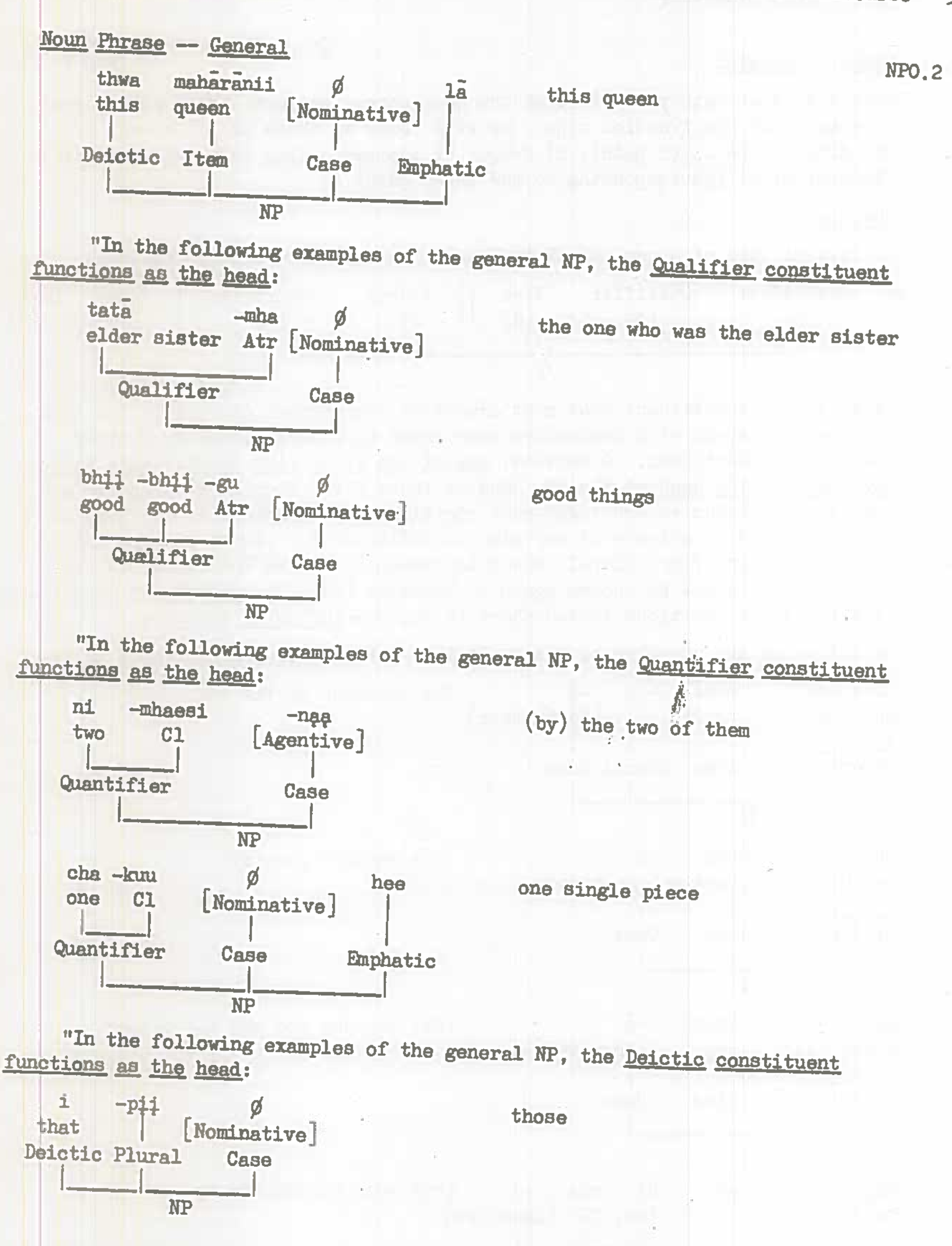
"The maximal form of a general NP in Newari may be represented as follows:



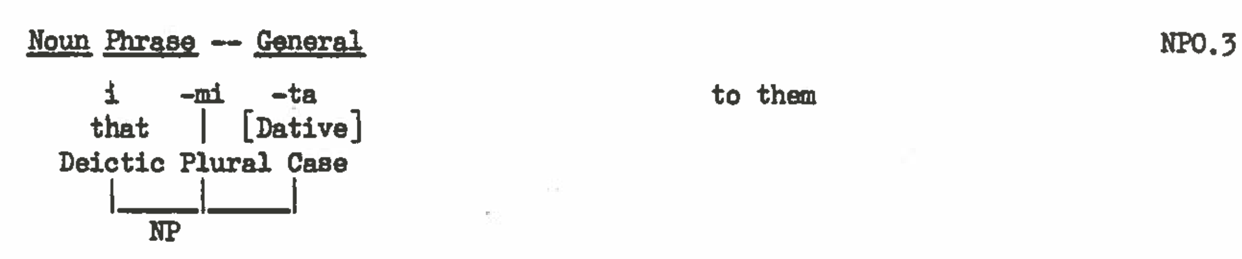
There is no single constituent that must always be represented overtly in the noun phrase. If we may speak of a nominative case which is always marked by Ø, then case is always a constituent. Otherwise, any of the first four constituents listed above may serve as the head of the NP. One of these first four must be chosen and which ever of them occupies the right-most position then functions as the head of the NP. If the head is animate it may also be inflected for plural (except that if the head is a quantifier, Plural cannot be chosen). If the Quantifier is chosen in position (2) it cannot be chosen again in position (6). The right-most constituent within the first six positions listed above is the one inflected for case.

"The following are examples of the general NP with the Item constituent as head:




Noun Phrase -- General
NPO.2
thwa maharanii Ø lā
this queen [Nominative]
Deictic Item Case Emphatic
NP
this queen
"In the following examples of the general NP, the Qualifier constituent functions as the head:
tatā -mha Ø
elder sister Atr [Nominative]
Qualifier Case
NP
the one who was the elder sister
bhịị -bhịị -gu Ø
good good Atr [Nominative]
Qualifier Case
NP
good things
"In the following examples of the general NP, the Quantifier constituent functions as the head:
ni -mhaesi -nạạ
two Cl [Agentive]
Quantifier Case
NP
(by) the two of them
cha -kuu Ø hee
one Cl [Nominative]
Quantifier Case Emphatic
NP
one single piece
"In the following examples of the general NP, the Deictic constituent functions as the head:
i -pịị Ø
that [Nominative]
Deictic Plural Case
NP
those

Noun Phrase -- General NPO.3



In this description, Deictic has not been treated as the head of a general NP but rather as the head of a pronominal NP. At this point the distinction seems fairly arbitrary.

"Within the first six positions of the NP, the right-most constituent is the one inflected for case. This can be seen from the following examples:

```
dee      cha-gul    -ii  [nasality of ii]   throughout the whole country
country  one Cl  [Locative]
|        |___|       |            |
Item   Quantifier  Case       Emphatic
|__________|_________|____________|
                NP

macā    -ta      -eta                        to the children
child    |     [Dative]
|        |        |
Item  Plural     Case
|________|________|
        NP

wa       khicā      -ā                       (by) that dog
that     dog    [Agentive]
|        |          |
Deictic  Item      Case
|________|__________|
        NP

tatā          -mhaes    -ita                 to the one who is elder sister
elder sister   Atr    [Dative]
|______________|         |
   Qualifier            Case
        |________________|
               NP

wa       yāta                                to that one
that   [Dative]
|        |
Deictic  Case
|________|
    NP
```

"There appear to be few restrictions on the independent choice of constituents within NP. There are few examples of a plural occurring with a post-head quantifier within the same NP. The following is from Prem Bahadur Kansakar, 'The children of the dog'.

<u>Noun Phrase</u> -- <u>General</u>

NPO.4

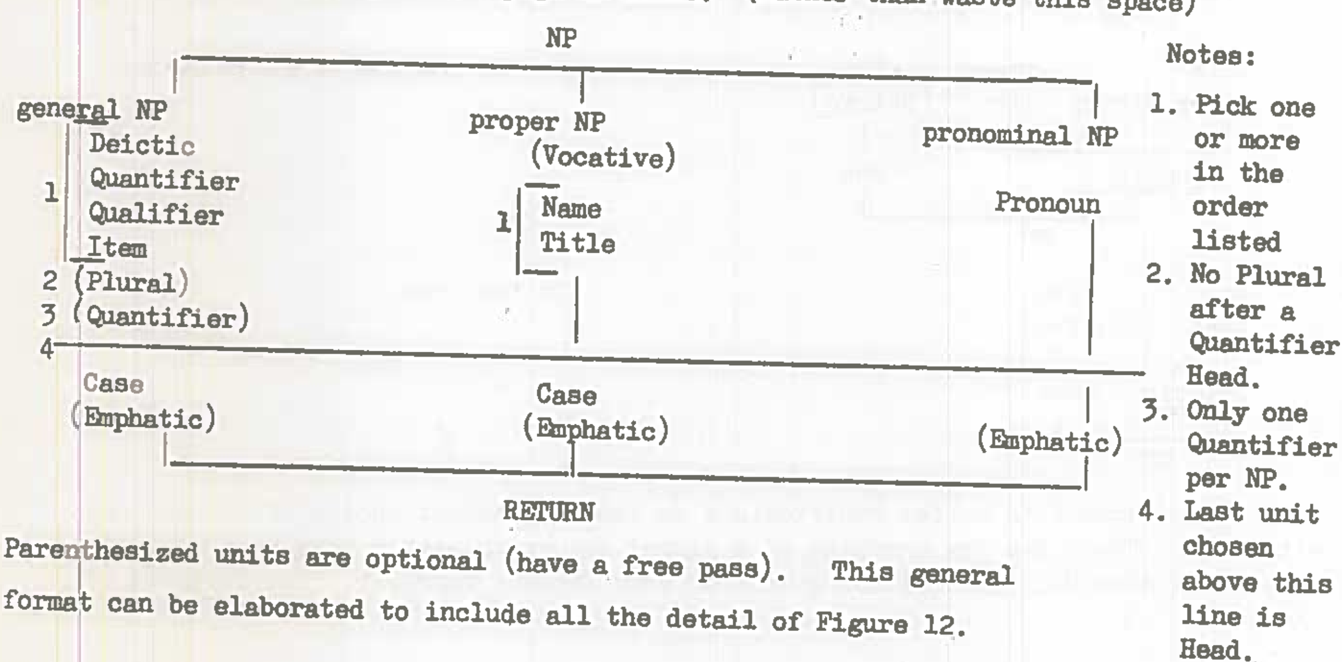
wa	macā	-ta	ni	-mha	ø	Those two children (6.14)
that	child	Plural	two	Cl	[Nominative]	

macā	-ta	ni	-mha	-a	(by) the two children (7.5)
child	Plural	two	Cl	[Agentive]	

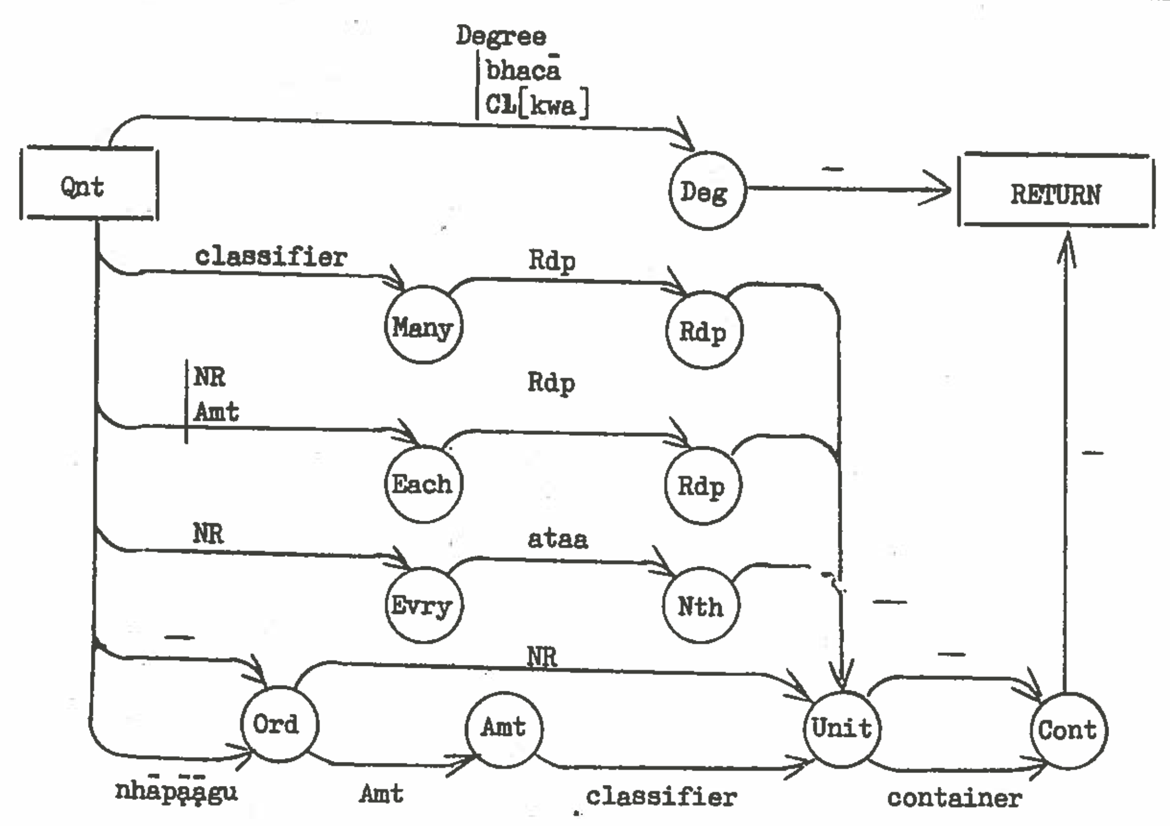
Plural, of course, is marked only in animate noun phrases, that is, in noun phrases in which the head is interpreted as referring to an animate being. In addition, plural cannot occur with Quantifier-head noun phrases. We have as yet no examples of a noun phrase with post-head quantifier in any noun phrase in which the head is not an Item."

As time goes by this section, Noun Phrase -- General should grow. Any patterns or regularities which concern the noun phrase in Newari but which involve two or more of the grammatical constituents in regular interaction will be recorded in this section. Any regularities that relate to a single constituent of the noun phrase will be discussed in the sub-sections that follow.

We pass on now to illustrate the second recommendation regarding the outline, namely, that subsections which have anything more than single-word structures to deal with be introduced in the same way as main sections, that is, with a network, footnotes, and a subsection outline. Since this will require a whole page and since many people find networks difficult to draw, we present below a simplified version of the network given on page 48 above. (Rather than waste this space)



Parenthesized units are optional (have a free pass). This general format can be elaborated to include all the detail of Figure 12.



Newari Noun Phrase Quantifier Network, 22 August 1978

Degree: bhacā, yekkwa, phukka, guli,

Amt: gwaa, gapāe, guli, taa, cii, ciki, cicā, cikicā, bharae, thapāe-cā, apāe-cā, sasipāe,

Section Outline

A. NR (numerical quantification)
- 1) Inflection 2.1
- 2) Classification 2.3
  - a) True classifiers
    - syntactic 2.3
    - unique 2.7
    - reduplicative 2.8
  - b) Direct quantified 2.9
  - c) Measure units 2.12
  - d) Containers 2.15
  - e) Quasi-units 2.20
  - f) Non-units 2.25

B. AMT (non-numerical quantification)
- 1) Intensives 2.30
- 2) Amount quantifiers 2.33
- 3) CL[kwa] comparative 2.35
- 4) Degree 2.36

C. Nr (numerals) 2.40

There are two basic kinds of quantifiers within the noun phrase: numerical and non-numerical. The basic function of the classifier system for numerical quantification is to provide a mechanism for counting otherwise uncountable entities. Countability presupposes CLOSURE. Closure is achieved grammatically with true classifiers, but lexically with other units. The basic function of non-numerical quantifiers is to specify amounts without counting or individuating the quantified mass. Classifiers can be viewed as Newari's means of deriving count nouns from mass nouns. Inherent count nouns are rare in Newari. Unquantified item noun in Newari are indeterminate for number. Classifiers are selected both by head noun and by numeral set. The ek set consists of Nepali numerals and goes primarily with measure units (chaa inci, sāth manā). The cha and chi sets are both Newari.

A Glossary of Technical Terms for Readers of Joe Grimes' Transition Network Grammars: A Guide (in Grimes, (Ed.) Network Grammars (SILP Nr. 45, Norman, Oklahoma, 1975) Ch. 3.

Larry L. Seaward

[Page number references are to Grimes, 1975.]

ABORT is an action that causes any arc on which it appears to fail even though the tests on it may have been met, the match found, and some actions performed. This forces the automatic parser to back up and try another arc, presumably one that will succeed on the basis of information tested on the way to the ABORT. P. 67

actions are taken when a match is found. An arc can have any number of actions associated with it, to be performed if the match works and ignored if it does not. Each specific action is enclosed in a pair of parentheses. Pp. 53-54.

ADDL puts the form specified at the left end of the contents of a register. P. 64

ADDR puts the form specified at the right end of the contents of a register. P. 64

AND a logic component, true only if all tests within its scope are true. P. 62

arc An arc is the route from one state to another. Each arc is described by giving its component parts. The parts of an arc are a match, the test, the actions and the terminal actions. The basic grammar uses only the match and the terminal action. Kinds of arcs are: DO, CAT, JUMP, MEM, POP, PUSH, TST, VIR, WRD. Pp. 47-51, 53, 57-59

ARC SET An arc set consists of the name of a state followed by the descriptions of all the arcs that leave that state, given in the order in which they are to be tried. P. 51

* precedes comments interspersed in the grammar or the dictionary. P. 81

@ says that during BUILDQ the different lists indicated (as X1 X2 ...) are to be appended to each other to form a single list. P. 66

atoms are elements of a list in LISP and are single words separated by spaces. They may be any string of characters that contains no spaces and that begins with an alphabetical character or a number with or without a sign. P. 66, 191

BUILD An action in which the skeleton, instead of being quoted directly as in BUILDQ, comes from the evaluation of some other expression. This might be used if there were several possible skeletons, the choice of which involved a COND conditional expression. P. 66

BUILDQ is an action used most often with POP arcs but may be used with other kinds as well. It has the form (BUILDQ SKELETON REGISTERS). Pp. 54, 56, 66-67

CAT stands for a category match. Pp. 49, 51, 58, 61

CATEGORY stands for a word category like DET (determiner) as used in the dictionary. P. 51

CHECKF is similar to GETF, except that it looks up the current word to see if it has the feature that is requested under the category that is named. P. 61

comment can be interspersed throughout the grammar or the dictionary as needed, having the form (* ...). P. 81

compound pair consists of the word COMPOUNDS followed by a list of <u>compound trees</u> enclosed in parentheses. P. 80

compound tree consists of a word, a result, and if necessary another tree, with a pair of parentheses enclosing the three. P. 80

COND is a conditional expression. These are alternate actions possible on some arcs, depending on specific details of what was found during a match. COND is followed by pairs of expressions and actions. Each expression, like a test on an arc, can be either true or false. Each pair is tried in turn. When an expression is found true, that action is taken and the condition being satisfied, no other pairs are looked at. Pp. 65, 66

COPY An action used to transfer the dictionary information about a word rather than the word itself. P. 68

current form is either the word that was matched by a CAT match, or else the entire structural description that was built up at a lower level by a POP arc. Current form is represented by ". P. 53

definiens The word entry in the dictionary that is being defined. P. 79

DETBUILD used for building complete sentences. Basically a BUILDQ action, but it is complex and may be called by more than one arc. Pp. 66-67

dictionary is where all the information that does not have a place in the grammar is kept. The LISP form of a dictionary entry consists of two parts: the <u>definiens</u> or word being defined and a list, enclosed in parentheses, of information <u>pairs</u>, giving the form (DEFINIENS (PAIRLISTS)). The first member of each pair is a key to the kind of information contained in the second member. Pp. 77-81

DO An arc allowing actions to be carried out unconditionally before transfer is made to the next state. P. 59

$ indicates an expression which during BUILDQ is to be evaluated and the results put in the place of the dollar sign. P. 66

EQ is a test which is followed by two things to be equated. P. 60

FEATURE A feature of a word as listed in the dictionary. P. 60

feature pair consists of the word FEATURES followed by a list of the feature names. P. 80

flag is a register which has been set to an arbitrary value such as T or NIL. P. 53

GETF gets the value of a feature from the dictionary. It is used only with CAT matches because it always refers to the current input word. P. 60

GETR makes the contents of the register named available. It can be used as a test: GETR is true if the register contains anything and false if empty. P. 60

GETROOT has as its value the root given by the dictionary for the word indicated by SOURCE with respect to the category named. P. 68

HOLD An action placing on a holding list for later removal and placement by a VIR (virtual arc), those constituents found out of their regular place in a construction. Pp. 53, 64

input text The text to be tested by the program, in this case, the network grammar.

INTRANS is a test that it true if the register contains a verb that cannot take an object. P. 61

JUMP A terminal action which causes the next state to examine the current word again instead of moving ahead to the next word. Pp. 54, 58-59

JUMP arc names a state to which a transition is to be made without advancing to the next input word, a test, and a list of actions. Pp. 54, 58-59

lexical category pairs can have four kinds of values paired with the category symbol: the morphology codes, the instruction for current form to direct morphological analyzer to put the text form itself out as the word that is recognized, the root feature list, and a list of root feature lists enclosed in parentheses, each giving a different interpretation of the word. Pp. 79-81

lexical category symbol is the first member of a lexical category pair, such as V (verb) or N (noun). P. 79

LIFTR sets a register on a higher level of embedding. It is used in reporting back the structure found by a PUSH arc. Pp. 63-64

LIFTRQ is the same as LIFTR, but eliminates QUOTE. Pp. 63-64

LISP is an artificial language that is suitable for the precise expression of functions that apply to complicated data structures. P. 191

list In LISP notation a list is represented by zero or more elements enclosed in parentheses. The elements of a list are either other lists or atoms which are single words separated by spaces. P. 191

logic Component of a test: NOT, OR, AND, NOR. P. 62

match seeks to make the match so named. If a match is found then one or more actions follow; if a match is not found then another arc is tried. Kinds of matches are: CAT by word category like DET or V used in the dictionary, PUSH state name like NP on an arc, WRD or MEM matching specific words or members of lists of words, VIR examines a HOLD list. Pp. 51, 59

MEM arc seeks to match input with one item of a following list of words. P. 58

MODAL is a test which is true if the current word is a modal like can, ought, or might. P. 61

morphology code The second member of a pair of which the first member is a lexical category symbol, such as IRR for irregular verb. P. 59

NEAREST The first occurance of the register named on a higher level; a value for WHERE. P. 60

NEXTWRD by itself makes available the word after the current one. With an argument NEXTWRD acts as a test condition. P. 60

NIL false test result. P. 53

NOR A logic component of a test, true only if all conditions within its scope are false. P. 62

NOT A logic component of a test. (NOT(test)) is true if the test is false and false if the test is true. P. 62

NPBUILD is used for building noun phrases. Basically a BUILDQ action, but it is complex and may be called by more than one arc. Pp. 66-67

NP FEATURES A register. Pp. 64-67

NULL EXP tests any LISP espression to see if it is false. P. 60

NULLR A test. The opposite of GETR used as a test. P. 60

OR A logic component of a test. (OR(test 1)(test2)) is true if any test of its scope is true. P. 62

pair The first member of a pair usually is the name of the type of information contained in the second member. Kinds of pairs are: compound pair, feature pair, particle pair, and substitute pair. P. 79

particle pair consists of the word PARTICLES followed by a list which is composed of pairs: a verb particle and the artificial definiens to be substituted for the verb if the combination is found. P. 80

PNCODE is a register (person-number category). Pp. 61, 67

PNCHECK tests the noun phrase named to see if it is of the person-number category required by PNCODE. P. 61

POP takes the data off of a pushdown stack; i.e. the data pops back up to the level of analysis that did the PUSH. POP is the equivalent of RETURN to calling routine. Pp. 48-49, 65-66

POP arc collects all the bits and pieces of information about a construction that were put into various registers since that part of the network was activated and places them into a representation of the construction that has been found. In building this representation it can change the order of elements and can add or subtract information. The structure that is built up by the POP is passed on to the arc that initiated the PUSH, which from there on handles it as a unit. Pp. 48-49, 53-54, 59, 65-66

PUSH implies pushing data down onto a stack (register). Pp. 49, (50-51), 52

PUSH arc An arc which makes a match by recursion to a different level of the network. When a match is made a POP arc brings the analysis to the state which was named by the PUSH arc. P. 63

pushdown stack A register for storing a list of items. Each new item occupies the first location in the register, all other items moving down one place in the register. Works on the principle: last in, first out. P. 49

Q Question

" stands for current form. If for a PUSH arc it stands for the entire construction that was matched. Pp. 60, 63

QUOTE indicates a character string to be put into a register instead of the current form. QUOTE is followed by the string to be entered as: (QUOTE DCL). Pp. 53-54, 60

QSTART is true if the current word is an auxiliary verb or one of the interrogative words that can begin a question. Pp. 59, 61

RESUME continues the analysis of a network where it was left off by RESUMETAG. Pp. 64-65

RESUMETAG sets a register in the construction that called the NP network if the construction is only partly analyzed when a POP is reached. P. 64

RFEAT is true if the word indicated has the feature named in its dictionary entry. P. 61

REGISTER is the name of some register, a pushdown stack. Registers are like notes that are kept about items or constructions that match arcs. For example: a register named SUBJ might be used to contain the noun phrase found on an arc which calls for NP filling subject slot. P. 52

REVERSE LIST has as its value a list derived from the one named but with its elements in reverse order. P. 68

root feature list The second member of a pair of which the first member is a lexical category symbol, such as PAST. The first part of the root feature list gives the root under which the rest of the information for this word is stored, while the second part gives a list of inflectional features. P. 79

S The starting state. P. 48

SBUILD is used for building complete sentences. Basically a BUILDQ action, but it is complex and may be called by more than one arc. Pp. 66-67

SCOMP is true if the register named contains a verb that is capable of taking a sentence complement, like want (to go) or feel (that it is time). P. 61

SENDR sets a register on the next level down rather than on the current level. This is done in preparation for a PUSH so that the information sent will be available to the network at the lower level. P. 63

SENDRQ substitutes the letter Q for the (QUOTE) in a SENDR action. P. 63

SETQ An action that assings a value to a variable that does not go into a register list, and therefore never enters the pushdown stack. P. 68

SETR takes the word or construction indicated and puts it into the register named. P. 53

SETRQ substitutes the letter Q for (QUOTE) in a SETR action. P. 63

SKELETON is an empty structure that contains plus signs to indicate the points where information is to be filled in. Following the skeleton come the names of registers where information is kept. The respective registers' contents replace the plus signs. Example construction: (POP (BUILDQ (S + + +) TYPE SUBJ AUX VP) T)). Pp. 54, 66

source of information is normally either a register, the input text, or the dictionary. P. 60

STATE A condition from which an arc, or arcs, parse(s) the grammatical structure connecting it to another state. (Woods' form of state labels, P. 58) P. 51

substitute pair does the opposite of what a compound pair does. It takes a single form like an abbreviation and replaces it with a string of words. It consists of the word SUBSTITUTE followed by a list of words to be put into the input string in place of the one that was found. Pp. 80-81

SUSPEND W suspends one computation in favor of another. It adds the weight W to the arc it is part of and puts that arc on the list of alternatives to be tried later if the next arcs fail. P. 67

T True; top-most level. Pp. 53, 60, 65

test Tests on arcs look for agreements; i.e., is the verb known to be of the kind that can take a noun phrase in that position? Every arc in the grammar has a test on it. Arcs that really require no test are given the test result T for "true" to show that the match is to be attempted under all circumstances. Tests other than the universal T are built up of three components: sources of information to be tested, test conditions to be applied to that information, and in the case of complex conditions, a logic for combining several tests into one. Pp. 52-53, 59-60

terminal action is (TO STATE) in which STATE names the state to which transition is made if the match succeeds. TO terminal action causes its following state to examine the next word in the input sequence. Or, (JUMP STATE) which makes the next state examine the current word instead of moving ahdea to the next word. Pp. 51, 54

TO is a terminal action giving the name of the state to which a transition is to be made after the action is performed if the match is successful. Pp. 51, 54, 58

TST arc allows a specialized test to be performed which may then lead to a transition or specific actions without affecting the input text. P. 59

VIR arc examines a special list called HOLD to see whether an element analyzed previously by a PUSH to STATE but recognized as being out of its deep structure position when it was found can be moved into the arc at that point. Pp. 59, 65

VPARTICLE (VPARTICLE (REGISTER PARTICLE)) is true if the word in the register named is a verb that can take the verb particle specified. (VPARTICLE REGISTER) is true if the register contains a verb that can take any verb particle. P. 61

VPASSIVE is true if the register named contains a verb that can be passivized. P. 61

VTRANS is true if the register named contains a verb that can take an object. P. 61

WHERE is a second argument for GETR which allows the contents of a register on a higher level to be examined. WHERE can have the value T for the topmost level, NEAREST for the first occurrence of the register named on a higher level, or a number for a particular stack level. P. 60

WRD is a word match. Pp. 58, 61